THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume XXVIII November, 1937 Number 8

INTERESTS OF PUBLIC SECONDARY-SCHOOL BOYS

J. W. M. ROTHNEY

Psycho-Educational Clinic, Harvard University

If interest is the key concept in contemporary education, as one writer has recently stated[1] and as countless others have implied in their writings, it would seem advisable, if not absolutely imperative, that we know what interests are likely to be exhibited in the behavior of groups of subjects who are to be subjected to educational procedures. In order to know in which places interest is likely to be exhibited, it is essential that we observe behavior of individuals, record our observations, combine them into some sort of measuring device, and evaluate their significance in relation to alleged outcomes of the educational process.

The need for recording, scaling, and evaluating interests becomes apparent to critical readers of educational literature, for the word "interest" has many connotations derived by various authors who use it to justify certain existing educational procedures and by others who wish to promulgate some new educational device or theory.[2] It would be difficult to get a definition of the word "interest" that would be satisfactory to any large group of educators, yet it would be equally hard to get any discussion of educational practices in which "interest" was not considered.

In the experiment to be reported herewith we have defined interests as behavior exhibited by students in situations where no apparent compulsion was being applied to make them exhibit such behavior.

[1] Hartmann, G. W.: *Educational Psychology.* C. E. Skinner (Ed.), Prentice-Hall, 1936, Chapter IV.

[2] The literature is much too lengthy to be considered here. Summaries of sections of the interest-achievement arguments are found in D. Fryer, *Measurement of Interests*, Henry Holt, 1932; in most discussions of Progressive Education movements; and in fact almost any recent treatise on educational methods and curricula construction.

We have secured, by a "diary technique," lists of interests (defined as above) from a large group of public secondary-school boys, have grouped and scaled them, and have attempted to determine their relationship to certain educational outcomes. We have, further, presented in full the lists obtained by our technique so that individuals may interpret them from several points of view. Before presenting the results of our experiment, however, we shall make a brief statement of the results of selected pertinent investigations on the subject of interest measurement and the relationship between interest and academic achievement.

A summary of typical opinions of various investigators concerning the relation of interest to academic achievement can be made as follows: Kelley (1914) concluded that interest was important in academic achievement but that it was only half as important as industry. Franklin (1924) concluded that interest was a potent factor in aptitude for school success. Wyman (1925) concluded that a child must have high intellectual interests in order to succeed. Cowdery (1926) concluded that academic standing had little or no relation to the number of interests in a professional field. Columba (1926) reported that the relationship between achievement in a school subject and preference for it was negligible. Kornhauser (1927) concluded that extreme interest feelings were not correlated with academic success. Commius and Shank (1927) found that the interest-achievement relationship varies in different school subjects. Jacobsen (1928) concluded that interests were only slightly predictive of educational achievement. Brainard (1928) concluded that academic standing in general or specific subjects had little relation to interests scores in the engineering field. King (1929) concluded that interest tests were inferior to mental tests, but showed definite value for prognosis of high-school achievement. Shuttleworth (1929) concluded that his interest inventory measured very much what was better measured by a college entrance examination. Garretson (1930) concluded that interests in curricular fields in the high school and abilities in these fields showed a negligible correlation. Langlie (1930) concluded that there are exceptional chances that students interested in certain subjects will make their best grades in those subjects. Bowman (1932) concluded that there was a substantial relationship between interest in, and performance of, a task within a narrow field of subject-matter. Segel (1933) concluded that scores obtained on the Strong Vocational Interest Blank contribute enough to the prediction of school achievement in general, and in specific sub-

jects, that its use for this purpose is warranted. Dunlap (1936) found a correlation of .50 between expressed preferences or interests for a particular school subject and success in that subject.

Examination of the summary of opinions of the various investigations concerned with the determination of the relation of interest to achievement reveals that there is considerable divergence of opinion. Wyman, Dunlap, and Franklin find a fairly high relationship between interest and achievement. Cowdery and Garretson find little or no relationship. The median position is taken by Jacobsen, who finds that interests are only slightly predictive of academic achievement.

A conservative estimate based on a study of the reports of the investigators of the interest-achievement relationship would place the correlation between these two variables at .10. The efficiency in forecasting achievement of the interest tests appears to be less than two per cent. The correlation coefficients and forecasting efficiencies vary widely, however, with the subjects of the experiments, and the methods used in measuring interests. Criticisms of experimental procedure follow.

Unsatisfactory methods of work and unsatisfactory interpretations of results are to be observed in several of the experiments described and summarized in the report immediately preceeding. It must be pointed out, however, that the interest measurement movement is less than twenty years old and that the shortcomings pointed out here are errors of pioneers who have been blazing the trail.

The main shortcomings in previous experiments may be listed as follows:

1. Failure to give consideration to the factors of the scholastic aptitude of the subjects of the experiments.
2. Failure to give consideration to the factor of chronological ages of the subjects.
3. Failure to use adequate statistical procedures.
4. Failure to give due consideration to the selection of items of interest inventories.
5. Failure to determine the effectiveness of interest measuring instruments on subjects other than those upon whom scoring keys were devised.

These five shortcomings are discussed briefly in the following pages.

1. The relationship between interest-test scores and intelligence-test scores is shown in the work of Shuttleworth (1929) who obtained a coefficient of correlation of .54 between interest-test scores and school

marks. When intelligence was held constant, this coefficient of .54 dropped to .29. Had Shuttleworth not performed the second operation and reported only that the correlation coefficient between interest and achievement was .54, his results would have been quite misleading.

There are numerous examples of results that are vitiated by failure to secure measures of scholastic aptitude of the experimental subjects, either by means of a reliable intelligence or achievement test. The work of Fryer (1927), Kornhauser (1927), Brainard (1928), Langlie (1930), Bowman (1932), and that part of the work of Garretson (1930) that deals with school grades, is of little value in describing the relationship between interests and school achievement, as measured by school marks, because no evidence has been presented to show the extent to which the achievement scores and interest scores were influenced by the academic ability of the subjects.

Considerable evidence has been obtained by Terman (1925) and others which shows that all objective tests of personality correlate to a slight degree because of the fact that each reaction of the individual is the total response of an organism. Fryer (1931) has compiled a number of reports concerning the relation between interest scores and intelligence-test scores. He finds that there is a positive but small relationship.

The value of interest tests must be judged by the extent to which they supplement, not merely overlap, the measure of ability. Failure to realize that, in so far as tests correlate with each other, they are measuring the same functions, and failure to determine the extent of overlapping of tests has resulted in misinterpretation of the results of many interest-achievement studies, particularly those reported immediately above.

2. In no study of the relationship between interests and achievement, so far reported, is due consideration given to the known relationship between chronological age and scholastic achievement. It appears that by failing to give consideration to this factor, investigators have ignored a small, but important, influence on achievement. There is not a high relationship between age and achievement in any school grade (the median of 60 correlation coefficients found by this writer was $-.19$ and the range was from $-.49$ to $+.15$ in a single grade level), but it is high enough, and common enough to warrant consideration.

3. Adequate statistical procedures are difficult to define at this stage of the interest measurement movement, for "adequateness"

has not yet been determined. We do know, however, that if small populations are used, the differences between groups must be large to be significant, and that populations of fifteen and twenty-five (Kornhauser 1927, and Hepner 1931) in validation groups are too small to yield dependable results.

The high bi-serial correlation coefficients reported by Garretson (1930) are due in large part to the bi-modal characteristics of his curricular groups, within and without. Brainard's (1928) arbitrary fixing of values for the responses to his inventory was unnecessary when more scientific weighting procedures were available. Rating methods such as those used by Fryer (1927) predetermine high correlations. Errors such as these are so common in the field of interest measurement that contradictory results have been obtained.

4. Examination of existing interest inventories reveals that little consideration has been given to the selection of items included in them. Investigators usually select items which they think are of interest to students or ask teachers to list what, in their opinions, are items of student interest. The first procedure is almost certain to result in biased selection or inadequate sampling of student interests. The second results in over-emphasis of certain kinds of what are considered to be desirable activities, and in a consequent lack of emphasis on activities in which students themselves profess to be interested.

The limitations imposed by the use of adult judgments of student interests are observed when comparison is made between the items of an interest inventory which the writer has developed in connection with the experiment to be reported herewith, and inventories previously constructed. The items of this inventory were secured from students of the experiment who were asked to keep a diary of the activities which they had carried on "just because they wanted to." The items differ significantly from the items contained in the inventories of Brainard (1928), Garretson (1930), and Hepner (1931), which were obtained from teachers and selected by the authors of those inventories.

The seriousness of the sampling error that is involved in the selection of interest inventory items has never been ascertained. In the past the selection has been haphazard, subject to whim and bias, and subject to over-emphasis of certain interests and under-emphasis of others. No satisfactory criteria have been developed up to the present time by which the selection of interest inventory items may be made. In the experiment to be reported herewith, a criterion of selection which has not hitherto been used is employed. These criteria are discussed at some length later in this paper.

5. Failure to determine the effectiveness of interest measuring instruments by administering the tests to groups of students other than those upon whom the scoring keys were devised constitutes a serious error. The effect of this serious error is magnified when tests, and results of the use of such tests, receive widespread publication.

Scores of the members of the groups on which interest scoring keys are devised are interesting, but meaningless as far as general application is concerned. Failure to check the value of the interest measuring devices has resulted in serious misinterpretation of results of interest-achievement experiments, and is, in large part, responsible for much of the unwarranted dependence that is placed on interest test scores.

As in the measurement of abilities movement, many insufficiently standardized and otherwise inadequate tests have been published. Hepner (1931), for example, published a test which proposed to measure educational and vocational interests. Instructions were given in the accompanying manual to the effect that the scores might be used in a *practical manner;* at the same time, pleas were made for more data so that the usefulness of the test might be determined.

Of all the measures of interest that have been developed to date, only two show sufficient promise to warrant their use. The Strong Vocational Interest Blank (1927) which measures how nearly a man's interests coincide with those of average men successfully engaged in certain occupations, has been developed to the extent that reliable differentiation between occupational groups can be accomplished. Wyman's Free Association Test of Interests (1925) has been used successfully by its author, but no check on its usefulness in other situations has been reported, probably because it presents a difficult scoring problem.

The Strong Vocational Interest Blank has not been standardized for high-school boys, and the Wyman test is not applicable in its present form to high-school procedures. The Dunlap Academic Interest Blank (1936) designed for junior-high-school students shows much promise, but at the present time, since the correlation between academic preferences scores and achievement is only .50, it has little predictive value. Other interest tests, for the reasons given in this section, are so inadequate that they cannot be used profitably at the high-school level.

PART I

The interest inventory described in this paper was constructed especially for this experiment when an examination of other existing

inventories had revealed that none was satisfactory for the purposes of the experiment. The two chief objections to other inventories were concerned with the scoring procedures and the selection of items.

No inventory available at the time our experiment was begun was scored for achievement in school subjects except that of Kelley (1914) which was out of date. However, this difficulty might have been overcome by weighting the responses to another inventory by the procedure described below, had the items been considered satisfactory.

The selection of interest inventory items on the judgment of teachers and investigators, notably subject to error, may over-emphasize certain activities while ignoring others of considerable value. The extent of error in this procedure cannot be determined until there is a more complete examination of student interests. Estimates of student interests by adults may be a fairly good substitute for careful observation of student activities, but there appears to be a better one—that of having students report their own interests. In this way we can avoid the possibility of whim and bias on the part of the investigator, increase the possibility of covering the whole field of a student's interests, and secure items with which the student is familiar.

It was decided that our subjects should write their own items. Accordingly, three hundred six boys who composed the test-building group in this experiment were visited in classes of approximately thirty-five, three times during a period of one week. Each student spent three school periods during the week in the writer's presence, making a list of the activities he had carried on during the week "just because he wanted to."

The visits were so arranged that the activities of the week-end were included. The visits also extended over a period of three months beginning the latter part of September, so that the difficulties of seasonal variation were avoided in part, and so that periodic happenings, such as the publication of the school paper, examination periods, and issuance of report cards, were met with at some time during the visits. The school opening week was represented as well as the week before the Christmas vacation.

From the student responses some nine hundred activities, persons, and things in which these students had interested themselves were tabulated.[1] Two procedures for putting these items into the form of an inventory were suggested. The first was that of determining the

[1] Tabulation was done by simple response; that is, reading one magazine was considered a different item from reading another.

frequency of mention, selecting a critical frequency and leaving out all items below this critical frequency. The second was that of presenting all items to the students and giving due weight to the frequency in evaluating or "weighting" the responses to the items.

The first alternative was disposed of because of the difficulty of determining a satisfactory critical frequency and because one could not be sure that frequency of mention was a satisfactory criterion of inclusion. Certain activities are not carried on as often as others. Latin classes are held every day, but Latin examinations only four times a year; football games are held once a week, but practice is held every day. There was also the possibility that students may have forgotten certain activities.

The second alternative appeared to be a better one except for the fact that the presentation of an inventory of nine hundred items would take so much testing time that boredom and mental sets would have too great an opportunity to develop. The labor of weighting so many items would not be incidental by any means, and the time taken to score a test of nine hundred items might make its use impracticable. The best choice under the circumstances appeared to be that of grouping the items; this was done to form an interest inventory of three hundred twenty-one items. To what extent this grouping has affected the results, we cannot be sure. From comments of students, oral and written on the blanks, from the striking out of certain activities in a series and the emphasis of others by underlining, it is suggested that certain groupings were objected to, but the number of objections was not great.

The choice of the *L* (like), *I* (indifferent), *D* (dislike) response scheme was made after a study of other schemes. There is little evidence on which to base a decision, but best results have been obtained by use of the *LID* response scheme.[1] This scheme seems to offer the least difficulty in getting uniform interpretation of the symbols when good examples are chosen to illustrate the method. During the testing of four hundred seventy-one students in our experiment, no questions concerning interpretation of the symbols were asked after the directions were read.

The *U* (unknown) response was added to carry out our belief that students should not be required to respond to items with which they are not familiar. This *U* response was not weighted because its use by a

[1] The Strong Vocational Interest Blank is probably the best interest test so far developed. The *LID* response scheme is used in it.

small group would make the weight insignificant,[1] and its use by a large group would make it unsatisfactory since little evidence of value would seem to result from making students report concerning activities with which they are not familiar.

The items were grouped into fourteen divisions as follows:

ITEMS CONCERNING:

1. School subjects.
2. Church and Sunday School.
3. Creative activities—painting, writing poetry, etc.
4. Pets and collections.
5. Trips and visits.
6. Social relationships.
7. Family relationships.
8. Sports and games.
9. Radio, movies, lectures, plays.
10. Reading.
11. Extra-curricular activities.[2]
12. School activities, not generally classified as extra-curricular.
13. Anti-social activities (stealing, ringing false fire alarms, etc.).
14. Miscellaneous activities, unclassified.

These divisions were kept in the test, except in the case of anti-social activities which were distributed among the others in an attempt to avoid a set against a list of such items.

There is little evidence concerning the effect of the position of an item on the kind of response, but what there is, offered by Matthews (1929), indicates that it has little effect and the maximum change is not likely to be greater than seven per cent. We cannot be sure, however, that this reported change was due entirely to the position of the item. In arranging our inventory the items were distributed within the various divisions by chance. This procedure appeared to be best because of the lack of evidence concerning any better procedure and the lack of opportunity to experiment with this particular problem.

The general problem of weighting responses to an interest test is one of determining the extent to which the items differentiate certain groups, and the extent to which data on which the differentiation is based, are due to chance.

The particular problem in weighting the responses to the items of this inventory was one of differentiating achieving students from non-

[1] See weighting technique in the following pages.

[2] In the general use of this term—clubs, debating societies, school paper, etc.

achieving students, and determining to what extent the differences might be due to chance.

In order to secure a measure of interests that was not related to intelligence and age, the criteria of achievement used were the discrepancy scores (*D*-scores), the derivation of which is described later. By use of these criteria, it is possible to secure scores uncorrelated with chronological age, and intelligence as measured by the Terman Group Test of Mental Ability.

The weighting procedure[1] is an attempt to improve on the method used by Cowdery (1925) and Strong (1926). Both methods are modifications of the technique devised by Kelley (1923) and both start with the regression coefficient

$$\bar{Y} = \gamma_{12}\frac{\sigma_2}{\sigma_1}x \tag{1}$$

from which is derived

$$b_{12} = r_{12}\frac{\sigma_2}{\sigma_1} \tag{2}$$

In Strong's work the unreliability of the correlation coefficient (certain small correlation coefficients may be due to chance) is determined and the variance of it applied to the weight. By simplification the weight b becomes:

$$b = \frac{\phi}{\sigma_1(1 - \phi^2)} \tag{3}$$

where ϕ is the product moment correlation between 2-point distributions and

$$\sigma_1 = \sqrt{PaPb} \tag{4}$$

σ_1 is defined as the SD of a dichotomous distribution where Pa is the proportion choosing one response to an item and Pb is the proportion choosing all other possible responses to that item, and $Pa + Pb = 1$.

[We have used bi-serial

$$\phi = \frac{\sqrt{PaPb(Ma - Mb)}}{\sigma_2} \tag{5}$$

in which Pa and Pb are the proportions of choosers and non-choosers, respectively, and σ_2 is the SD of the criterion scores. Bi-serial ϕ

[1] I am indebted to Professors T. L. Kelley and P. J. Rulon for assistance with the statistical procedures described in this paper.

differs from bi-serial r in that it does not assume normal distribution of the dichotomized variable—a doubtful assumption concerning definite responses to inventory items.]

In our weighting procedure, in addition to changing to the use of bi-serial ϕ, we have determined and used the unreliability of the whole regression coefficient instead of using Strong's method which used only the unreliability of the correlation coefficient. By this procedure we hoped to minimize the influence of chance fluctuations which might occur in like sets of data.

The derivation of the formula for the standard error of a regression coefficient is provided by Kelley (1923, p. 176, formula 107). When the variance of the standard error of the whole regression coefficient is determined and applied to the regression coefficient the weight b' is obtained from

$$b' = \frac{r_{12} - \sigma_1}{1 - r_{12}} \tag{6}$$

Our r is bi-serial ϕ, and σ_1 is defined as above.

The formula for obtaining b' given above (6) is a naming formula. Substituting for r_{12}, σ_1 and σ_2, the identities given in formulas (4) and (5), and multiplying by 10 to eliminate decimals, we have the work formula:

$$b' = \frac{10(Ma - Mb)}{\frac{1}{PaPb} - \frac{(Ma - Mb)^2}{{\sigma_2}^2}} \tag{7}$$

A plus sign was attached to the weight of the response chosen by the group which had the highest mean criterion score, and a minus sign to the response chosen by the lowest criterion score group. The work formula was used in weighting responses of three hundred six students to each of three hundred twenty-one three-choice items in the inventory.

The weighting procedure for the response L to the item "geometry" is given below with a diagram descriptive of it.

Let Y = the continuous variable, that is, the geometry discrepancy scores.

Let X = the dichotomized variable, that is, those not choosing the response and those choosing the response.

Let Ma = the mean discrepancy score of those students who chose the L response.

Let Mb = the mean discrepancy score of those students who did not choose the response L.

Let na = the number of students who chose the response L.
Let nb = the number of students *not* choosing the response L
and $N = na + nb$

Then $$\sqrt{PaPb} = \sigma_1 = \sqrt{\frac{na}{N} \cdot \frac{nb}{N}}$$

and σ_2 = standard deviation of geometry discrepancy scores.

Substituting the numerical values as follows:

$$\begin{aligned} Mb - Ma &= +5.43 \\ N &= 304 \\ na &= 166 \\ nb &= 138 \\ \sigma_2 &= 7.45 \end{aligned}$$

the weight to be attached to a response L to item "geometry" on the geometry scale is +15.

The weighting procedure was applied to each response six times in order to devise keys for English, Latin, French, Geometry, Algebra, and average scholastic interest. The weights range from −26 to +15, but the majority lie between −5 and +5. The complete scoring keys are much too extensive to present here, but a general idea of the weights may be observed by examination of the weights attached to the responses L (like), I (indifferent), and D (dislike) to the item "geometry." These weights indicate that the marking of the L (like) response is done more frequently by the successful students in all subjects, and the marking of the I (indifferent) and D (dislike) responses is done by students who are less successful. The large weights that are attached to the L (like) response to the item "geometry" on the geometry scales indicate that the marking of this response is more indicative of success in that subject than the marking of either of the other responses.

	English	Latin	French	Geometry	Algebra	Average
Geometry L	5	3	0	15	4	5
Geometry I	−2	−3	1	− 5	−1	−2
Geometry D	−3	0	−1	−10	−3	−2

It should be pointed out here that no assumption is made concerning the truthfulness of the responses to these items. When it was found

that one hundred sixty-six out of three hundred four students who responded to the item "geometry" encircled the *L* response, and that the mean achievement scores of this group was five points higher than the group that encircled the *I* and *D* responses, it was possible to derive a weight that indicated the relationship between marking the *L* and geometry achievement. We cannot, however, conclude that the one hundred sixty-six students liked geometry. Many of the students may have disliked the subject, but may have set out deliberately to create a false impression. The faking of responses is an inherent limitation of the reaction questionnaire technique, but the limitation does not appear to be so great as to require the abandonment of it as a research technique. A clear distinction must be understood, however, between this kind of questionnaire and the more widely used "fact-finding" questionnaire.

The selection of subjects for our experiment was governed by the results of previous investigations, the necessity of securing a representative sampling of students of public high schools, and the availability of students who satisfied the requirements imposed by the first two criteria. The selection has been such that the results obtained should be applicable to students of public high schools generally. It is quite possible that a particular school might develop measurements which were applicable in that school only, but experiments such as the one reported here must be directed toward the discovery of results which are quite general in their application.

The limitation to one curricular group was made to avoid the complications of differences in interests of curricular groups suggested by the studies of Franklin (1924), and Garretson (1930). This problem of curricular interest differences is an important one that needs further investigation, but it was not thought advisable to investigate it in connection with this experiment. The choice of the college group was prompted somewhat by the standardization of requirement by the provisions of college entrance.

The accident of grade placement was not considered important enough to require the conduct of this experiment in any particular grade. However, most students have moved from the junior high school at the beginning of the tenth school year; on that account, the choice of their tenth-grade marks as criteria was made because the marks of students at this level are less likely to represent past performance. The fact that all students in this grade are required to take exactly the same subjects during the tenth school year also influenced

the choice of the tenth-grade marks because the opportunity was provided for the study of the achievement of students who were carrying the same course load.

The schools from which the students in this experiment were drawn have been rated in effectiveness as representative of the larger cities of the State of Massachusetts, but have been rated as slightly better than the larger high schools of the United States as a whole.

Eight of the ten schools from which students were selected are organized on the 6-3-3 plan, and two have retained the 8-4 plan. Except for this difference which, as Beatley (1932) points out, may or may not be of importance, there is no reason to believe that there are great differences in the previous school history of the students.

The students of the experiment resided in cities which compose a representative sampling of the socio-economic status of the cities of Massachusetts. The sampling is, however, slightly superior to the cities of the United States as a whole.

The studies of Terman (1925), Lehman and Witty (1927), Columba (1926), and a summary of foreign studies reported by Fryer (1931) suggest that the educational interests of the two sexes are sufficiently different to make it impossible to combine them for any practical use to which the results might be put. In view of this evidence it was considered advisable to select boys only as subjects for this experiment.

The chronological ages and intelligence-test scores of the subjects showed the usual variation found in a single grade. Examination of Table I reveals the similarity of the groups in intelligence (as measured by the Terman Group Test of Mental Ability) and in chronological age. The means of the groups of the various schools range from one hundred thirty-eight points in raw score to one hundred seventy-one points. This difference is equivalent to a difference of sixteen months of mental age, but the differences within any particular group are greater than the group differences. The greatest difference between groups in chronological age is seven months, but again there are large differences within the groups. Since age and intelligence are held constant, the slight variability noticed is not an adverse condition in this experiment.

The criteria of school success were the grades obtained by the students during the ordinary course of their school work over a period of one year. Examination of the grades of our groups of students revealed that there were large differences in the grading systems of the schools, and in the courses given in a particular school. An examina-

TABLE I.—MEANS AND STANDARD DEVIATIONS OF RAW SCORES OBTAINED ON THE TERMAN GROUP TEST OF MENTAL ABILITY, AND OF CHRONOLOGICAL AGE. STUDENTS IN THE ELEVENTH GRADE OF TEN HIGH SCHOOLS IN MASSACHUSETTS

School	*N*	Terman raw scores		Chronological age (months)	
		Mean	σ	Mean	σ
A	37	157.90	25.64	195.8	7.35
B	46	157.76	26.17	192.5	8.00
C	30	138.18	21.86	191.5	6.86
D	56	157.19	24.26	193.8	9.74
E	94	145.14	27.93	195.1	8.54
F	20	171.00	19.21	193.8	7.86
G	23	159.88	19.16	199.1	7.41
H	49	161.69	19.16	195.4	7.44
I	47	171.03	23.96	199.8	10.76
J	53	163.23	22.79	198.4	9.42

tion of Table II, in which are given the systems most commonly found in the schools, reveals the extent of these differences. It will be seen that the number of categories into which a student may be graded varies from four to sixty. All the letters from *A* to *F* are used, with various plus and minus signs attached. The percentage system is retained by two schools and one school has a five-mark division system extending from fifty to one hundred. It is quite impossible to indicate any average grading system that would describe adequately such a heterogeneous group of marking systems.

It is practically impossible for teachers in the ordinary course of their work to determine the achievement of their students as reliably as these grade classifications suggest.[1] There does appear to be a tendency, however, to give grades in four classifications corresponding to honor, college credit, pass, and fail, represented by the grades *A*, *B*, *C*, *D*, and these classifications were used except in cases where there was a definite five-grade classification or where only three grades had been given. The procedure used was that of grouping all *A* plus, *A*, and *A* minus, grades as *A*; all *B* plus, *B*, *B* minus grades, as *B*; etc.

The method of evaluating these grades in terms of standard scores is given by Kelley (1923, p. 101). The grades are expressed as devia-

[1] For example, it would be necessary to construct tests with reliabilities of .95 in order to place eighty per cent of the students correctly on a five-point (*A* to *E*) scale.

tions from the mean of the whole group by considering each classification as representing a proportion of a normal distribution. The mean of the whole group in terms of standard deviations is computed,

TABLE II.—THE GRADING SYSTEMS USED IN TEN SENIOR HIGH SCHOOLS OF MASSACHUSETTS

School	Grade categories	Number of categories
A	*A*, *A*−, *B*, *B*, *B*−, *C*, *C*, *C*−, *D*, *D*, *D*−	11
B	40 to 100	60
C	*A* *B* *C* *F*	4
D	*A*, *A*−, *B*, *B* *C* *D* *E*	7
E	*A*, *A*−, *B*, *B* *C* *D* *E*	7
F	*A* *B* *C*	4
G	*A*, *A*, *A*−, *B*, *B*, *B*− *C*, *C*− *D* *E*	10
H	*A* *B* *C* *D* *F*	5
I	95, 90, 85, 80, 75, 70, 65, 60, 55, 50	10
J	50 to 100	50

and the value is given to the classification. In our experiment the values were multiplied by ten and the nearest whole number was taken to eliminate decimal places in the computation. After the grades had been transmuted into standard scores, it was then necessary to substitute these scores for the letter grades, and divide by the number of such grades to get the yearly average in each school subject.

All the grades received by the students were translated into numerical scores on an equal-unit scale so that for each student we had a usable index of his achievements in five subjects, based upon the achievement of the groups in which he was working. In order to get criteria of success which were uninfluenced by the factors of chronological age and intelligence, one more step was necessary before the achievement scores could be correlated with test scores to determine the relationship between interests and achievement.

To eliminate the influence of scholastic aptitude and chronological age on school achievement the following procedure was utilized. Predictions were made of the students' achievement on the basis of scores on the Terman Group Test of Mental Ability and chronological age. Actual scores were then subtracted from predicted scores to obtain measures of achievement uncorrelated with age and intelligence test scores. The measures which resulted will be referred to hereafter as *D* (discrepancy) scores.

The correlation coefficients between the interest inventory scores and school achievement presented in Table III reveal the extent to which the weighting procedure described previously had selected achieving and non-achieving students in each school subject.

TABLE III.—CORRELATIONS BETWEEN SCORES ON THE INTEREST INVENTORY AND *D*-SCORES OF ACHIEVEMENT IN SCHOOL SUBJECTS. TEST-BUILDING GROUP OF THREE HUNDRED SIX BOYS IN SEVEN HIGH SCHOOLS IN MASSACHUSETTS

	1 Average interest	2 English interest	3 Latin interest	4 French interest	5 Geometry interest	6 Algebra interest
Average *D*-score.....	*.49*	.43	.45	.44	.45	.39
English *D*-score.....	.35	*.50*	.28	.27	.25	.24
Latin *D*-score.....	.38	.28	*.49*	.31	.28	.28
French *D*-score.....	.40	.31	.32	*.54*	.27	.34
Geometry *D*-score.....	.39	.27	.29	.27	*.58*	.36
Algebra *D*-score.....	.29	.22	.25	.27	.31	*.55*

The coefficients in the diagonal indicate that the relation between interest in a particular subject and achievement in that subject is correlated to the extent of .49 to .58. Average achievement, it appears, can be predicted with as much success as Latin achievement, and with nearly as much success as achievement in the other subjects.

The coefficients outside of the diagonal show that there is not a high degree of specialized subject interest. Coefficients ranging from .24 to .40 indicate the relation between interest in one subject and achievement in another.

It must be noted, however, that the coefficients of correlation between interest inventory and school achievement scores reported in Table III are computed from scores obtained by the group upon which the scoring keys were devised. It is necessary, therefore, to wait until scores have been obtained from an experimental group before the adequacy of the interest inventory is determined.

Since the items of the inventory used in this experiment were obtained directly from a large group of high-school students, there appeared to be less likelihood of the occurrence of wide variations in response of groups other than the test-building group. In addition due consideration was given in the weighting procedure to the possibil-

ity that there might be small chance fluctuations in the responses of other groups.

TABLE IV.—CORRELATIONS BETWEEN SCORES ON THE INTEREST INVENTORY AND SCORES OF SCHOOL ACHIEVEMENT. EXPERIMENTAL GROUP COMPOSED OF ONE HUNDRED FORTY-NINE BOYS IN THREE HIGH SCHOOLS IN MASSACHUSETTS

	1 Average interest	2 English interest	3 Latin interest	4 French interest	5 Geometry interest	6 Algebra interest
Average *D*-score.....	*.14*	.07	.14	.16	.08	.15
English *D*-score.....	.09	*.08*	.10	.06	.02	.09
Latin *D*-score.....	.18	.17	*.21*	.11	.08	.16
French *D*-score.....	.06	.03	.01	*.12*	.01	.04
Geometry *D*-score.....	.19	.09	.19	.17	*.21*	.20
Algebra *D*-score.....	.09	.01	.02	.14	.13	*.17*

The coefficients in the diagonal of Table IV which indicate the relationship between various school-subject achievement scores and the inventory scores descriptive of interest in those subjects, reveal the extent to which the inventory procedure has been successful.

It will be observed by examination of Table III that the diagonal coefficients indicating the relationship between school achievement and subject interests of the test-building group were substantially high (.49, .50, .49, .54, .58, and .55). They indicated that the test had differentiated between achieving and non-achieving students to a marked extent. The coefficients of correlation between the same variables in the experimental group are: .14, .08, .21, .12, .21, and .17. The size of these coefficients demonstrates that the test does not differentiate between achieving and non-achieving students on whom the scoring weights were not devised.

Our findings are similar to those of Kornhauser (1927) and Jacobson (1928) who were working at the college level.

It must be pointed out here that unusual care was taken in the construction of the inventory used in this experiment. Examination of the procedures described previously will reveal that due consideration was given to the chronological ages and intelligence (measured by the Terman Group Test on Mental Ability) of the students. The weighting of the inventory items was done by a procedure designed to mini-

mize the effect of chance fluctuations in like sets of data, and the items were expressions of students' voluntary activities.

With all these factors taken into consideration, the greatest yield in forecasting school success of any of the school-subject interest scales is less than two per cent. It appears that when such carefully devised instruments fail to secure results, there can be little doubt that the inventorying of students' interests for the purpose of predicting their scholastic success is not a profitable procedure. Achievement groups that may be differentiated by inventory procedures have not been found.

Further evidence of the failure of the inventory to differentiate groups is offered by the coefficients outside of the diagonal in Table IV. The very favorable situation noticed in Table III, where the coefficients in the diagonal are all considerably larger than those outside of the diagonal, is not present in the results obtained on the experimental group.

There is, for example, a higher correlation between average interest scores and geometry achievement scores than there is between average interest scores and average achievement scores; English interest scores are more closely related to Latin and Geometry achievement scores than they are to English achievement scores; French interest scores are more highly correlated with English, Geometry, and Algebra achievement scores than they are with French achievement scores.

Our findings indicate that the Average, English, and French inventory scoring keys measure interest in other subjects better than they measure the interests which they were built to measure. However, the standard errors of the coefficients are large and the variations well within their limits. Another administration of the test might result in securing coefficients which showed that the keys measure the subject interests that they were built to measure better than they measure any other subject interests.

The consistency with which small correlations between the test scores and the criteria of school successes were obtained led us to make an examination of the tests and the criteria in an effort to determine what factors were operating to cause these consistently small correlations. The investigation led first to a study of the reliability of the tests, for if they were not measuring consistently, this fact alone might explain the low correlations.

The reliabilities of the six scales of the interest inventory were determined by the split-half Spearman-Brown procedure. The coeffi-

cients were computed from scores obtained by students who were not in the group upon which the keys were devised. An examination of Table V reveals that the test is as reliable as the best tests of personality and many of the better tests of ability. Five of the six coefficients are greater than .80.

TABLE V.—RELIABILITY COEFFICIENTS, STANDARD DEVIATIONS OF SCORES, AND PROBABLE ERRORS OF THE RELIABILITY COEFFICIENTS OF THE INTEREST INVENTORY. ONE HUNDRED FORTY-NINE HIGH-SCHOOL STUDENTS

	r_{11}	σ	PE_r
Average interest	.86	39.29	.014
English interest	.81	35.06	.019
Latin interest	.80	41.49	.020
French interest	.85	46.09	.016
Geometry interest	.83	37.64	.018
Algebra interest	.79	45.38	.021

Granting the limitations imposed by the reaction questionnaire procedure, it is quite possible that the tests might be of some value in forecasting very good criteria. It is necessary, then, to examine the criteria to discover if any of the findings are due to unreliable or invalid school marks.

The reliability of the marks with which the test scores are correlated was obtained by the following procedure. From the four marks given to each student during the school year, three pairs were selected as follows:

First and second against third and fourth.
First and third against second and fourth.
First and fourth against second and third.

These pairs were thrown into one scatter diagram and the correlation computed. Since these marks are not independent, it was not possible to use the Spearman-Brown formula for finding the reliability of all the grades obtained. The reliability of the half grades (.83) used in the correcting for attenuation must be considered as slightly less than the reliability of all the grades.

Using the obtained reliability of half the grades and the reliability coefficients reported in Table V we corrected for attenuation the correlation coefficients reported in Tables III and IV. The corrections

raise the coefficients so slightly that they fail to indicate a close relationship between the subject interests and academic achievement.

A partial explanation of the low correlations is to be found in the effect of "halo" and "hearsay" in alloting of school grades. Symonds (1931) has summarized the evidence offered by a number of investigators concerning the effect of "halo." There is unanimous agreement among them that the "halo" effect is present unless raters have been taught to avoid it. The influence of the "halo" is carried on so that a student may receive high marks even after he has lost all interest in a subject and has stopped achieving in it. If we give him an interest inventory at this latter stage, he may show no interest in a subject, but his previous interest and achievement is remembered and his marks may be based on this memory.

Terman (1916) states that teachers err in judging intelligence due to their tendency to over-estimate the intelligence of the sprightly, talkative, sanguine child, and to under-estimate the intelligence of the child who is less emotional, who reacts slowly, and who talks little. It is quite possible that similar errors are made in the judging of achievement.

Such factors, quite outside of and beyond the control of the student, may operate to the extent that reliable differentiation of achievement groups, when achievement is measured by teachers' marks, may be impossible. Certainly these chance factors are not predictable by means of tests.

In summary, it may be stated that the relationships between our reaction questionnaire scores and school marks is not substantially greater than the small obtained coefficients indicate. The small relationships were not due to the unreliability of the tests nor of the criteria, but they appear to be due to the limitations imposed by the reaction questionnaire technique, and factors, unpredictable by tests, which influence the allotment of school marks.

The findings reported in this paper do not, however, indicate that there is no substantial relationship between interest in and achievement of a particular school task. It is quite possible that those phases or functions of personality represented by the behavior of an individual in reacting to stimulating school experiences drawn out of, or put into, material which is being taught by an excellent teacher, escape measurement by the reaction questionnaire procedure. The difference between the interest "observed" by the teacher and "expressed" by the student

appears to be so great that testing procedures have not been devised which make it possible to determine one from a knowledge of the other.[1]

PART II

Since we have found that our inventory has no value for the prediction of academic success, we may seek for values which it might have in the understanding of such subjects as we have chosen to study.

Among the values which have been attributed to interest inventory results are their value (1) in educational and vocational guidance, (2) in sectioning interest and social groups, (3) in giving individual assignments suited to individual interests, (4) in assisting the adjustment of problem children.

We have listed at great length the responses of each of our three hundred six subjects to each of the three hundred twenty-one items of the inventory. The reader should, before making a study of our items, review the statements concerning their origin, should note that these items were obtained from the students by means of the "diary" technique, and should note that they consist of a concentration of some nine hundred items. The responses indicated here are the responses to the items after they had been drawn up into inventory form and resubmitted for checking by the subjects. It will be well to review also the descriptions of our subjects.

The list of items has been presented in full so that examination of them may be made by readers who approach them from different points of view. Thus, those who are interested in students' reading preferences will find considerable interesting material. A student of the writer's, for example, finds that reading Shakespeare is ranked very low among the kinds of material which these people read. The student of delinquency will find no anti-social items in the upper fifty per cent of the preferred list of activities. Movie choices provide an interesting study for those who are interested in visual education. Stated interest in school subjects by students who have been taking the same subjects can be studied by those who are interested in students' choice of electives. Counsellors interested in preferences for work experiences may find some suitable information, and students of personality and mental hygiene can find some suggestions, in this tabulation, of contents of the mind. Those who are interested in the construction of interest inven-

[1] Since this was written the work of Jack Dunlap has been called to the writer's attention. His work would seem to offer more promise than any other work of its kind.

tories for diagnostic purposes will find a much needed "interest-frequency" list for the purpose of differentiating common from rare items of interest.

TABLE VI.—NUMBER OF ACTIVITIES LIKED BY STATED PERCENTAGES OF THREE HUNDRED SIX ELEVENTH-GRADE BOYS IN EIGHT PUBLIC HIGH SCHOOLS OF MASSACHUSETTS

Activities	91 to 100	81 to 90	71 to 80	61 to 70	51 to 60	41 to 50	31 to 40	21 to 30	11 to 20	0 to 10	Total
1. School subjects	..	..	..	3	6	2	1	1	2	2	17
2. Church and Sunday School	..	..	..	..	..	1	1	1	3	..	6
3. Creative activities, (painting, music, writing and poetry, etc.)	..	..	..	..	..	2	7	4	..	2	15
4. Pets and collections	..	..	..	..	1	2	1	1	1	..	6
5. Trips and visits	1	2	3	3	1	4	2	..	..	..	16
6. Social relationships	..	..	2	4	4	4	10	4	4	1	33
7. Family relationships	..	..	..	..	3	1	2	..	1	..	7
8. Sports and games	1	5	2	2	2	1	1	..	..	..	14
9. Radio, movies, lectures, and plays	1	1	9	5	4	2	4	3	1	..	30
10. Reading	..	2	3	8	9	13	6	7	6	1	55
11. Extra-curricular activities	..	1	..	..	1	3	4	5	3	1	18
12. School activities not generally classified as extra-curricular	..	3	..	3	2	2	3	..	3	..	16
13. Anti-social activities (stealing, ringing false fire alarms, etc.)	..	..	..	..	..	..	1	2	14	12	29
14. Miscellaneous activities, unclassified	..	..	2	8	8	12	6	11	10	2	59
Total	3	14	21	36	41	49	49	39	48	21	321

In order to make such studies more readily and in order to indicate the nature of the preferences of our subjects we have divided the items into ten groups on the basis of the frequency with which they were responded to by the marking of the "like" choice. Thus our first group consists of three items which at least ninety per cent of our subjects liked. The second group consists of fourteen items which more than eighty but less than ninety per cent of our group liked, the third group consists of twenty-one items which more than seventy but less than eighty per cent of our group liked. The other seven groups were sorted out in the same way.

TABLE VII.—NUMBERS OF SUBJECTS IN A GROUP OF THREE HUNDRED SIX, RESPONDING TO THE ITEMS LISTED. SUBJECTS WERE HIGH-SCHOOL BOYS IN EIGHT PUBLIC HIGH SCHOOLS IN MASSACHUSETTS

	L, like	*I*, indifferent	*D*, dislike	*U*, unknown or not experienced
Playing vigorous sports like swimming, skating, rowing	289	7	5	5
Visiting the beach	278	24	4	
Sport movies like "The Spirit of Notre Dame," "70,000 Witnesses"	275	24	7	
Graduation	274	15	5	12
Playing group games like foot-ball, baseball, hockey	270	23	13	
Visiting a camp	269	21	3	13
Honor-roll	267	21	3	15
Watching group games like foot-ball, baseball, hockey	266	31	9	
Listening to comedians like Ed Wynn, Amos and Andy	263	32	9	2
Diploma	258	29	3	16
Watching athletics like boxing, wrestling, track	257	31	12	6
Reading the news of my own town and school	254	39	9	4
Playing gymnasium games like basket-ball, volley-ball	251	35	20	
Other out-door activities like hunting, fishing, riding	251	24	7	24
Taking part in athletics like boxing, wrestling, track	249	35	16	6
Reading the sport page	248	46	12	
Gymnasium	247	39	17	3
News reels and travel talks	241	43	22	
Animal pictures like "Bring 'Em Back Alive" and "Trader Horn"	240	51	15	
Vaudeville	237	43	26	
Watching airplane maneuvers	235	48	14	9
Adventure movies like "Tiger Shark" and "The Blonde Captive"	234	38	17	17
Keeping my trousers pressed and my shoes shined	231	50	24	1
Playing games like tennis, ping-pong, hand-ball, croquet	230	39	25	12
Reading funny magazines like *Life*, *Judge*, *Punch*	230	53	11	12
Visiting colleges like Harvard and Tufts	229	21	1	55

TABLE VII.—*Continued*

	L, like	*I*, indifferent	*D*, dislike	*U*, unknown or not experienced
Trying for distance on the radio	229	52	14	11
Telling and listening to jokes	227	59	14	6
Watching for regular programs	227	53	18	8
Listening to radio dance music	226	54	26	
Swimming lessons	225	23	6	52
Reading scientific magazines like *Popular Science, Popular Mechanics*	225	50	24	7
Reading about or listening to reports of sporting events	223	44	31	8
Mickey Mouse and Silly Symphonies	220	43	31	12
Visiting museums like biological and art museums	219	50	17	20
Visiting historical places like battlefields, monuments	219	66	14	7
Buying presents for other members of the family	216	58	24	8
Going to plays and musical comedies	216	47	30	13
Men teachers	213	84	6	3
Listening to murder and mystery stories	213	53	36	4
Movies like "Horse Feathers" and "The Kid from Spain"	213	26	58	9
Planning my future after High School	212	56	26	12
Raiding the ice-box	211	49	10	36
Buying new clothes	211	62	26	7
Dreaming about my future life and work	210	60	21	15
Study periods	209	70	22	5
Sport books like *What Price Foot-ball, Base-ball News*	208	66	19	13
Reading the funnies and joke columns	207	32	61	6
Playing card games like whist, bridge, rummy, cribbage, casino	206	50	36	14
Musical comedy movies like "Love Me Tonight" and "Whoopee"	205	56	30	15
Walking and going to places like school with a gang	201	71	26	8
Reading news summary magazines like *Literary Digest, Time*	201	63	32	10
Sea Stories like *Gloucester Man, The Sea Hawk*	201	68	24	13
Saving my money in a bank or Christmas club	199	44	21	42
Visiting flower shows, food, poultry, and auto shows	197	51	26	32

TABLE VII.—*Continued*

	L, like	*I*, indifferent	*D*, dislike	*U*, unknown or not experienced
Arguing about many problems	197	62	35	12
Looking at boats and boat models	197	73	22	14
Listening to news and weather broadcasts	196	75	29	6
Algebra	195	59	52	0
Reading boys' magazines like *American Boy* and *Boys' Life*	195	75	26	10
History	194	83	24	5
Discussing radio programs and movies	194	75	24	13
Gang movies like "Scarface" and "Taxi"	194	74	29	9
Experimenting with chemical and electric things	193	31	21	61
Going to school parties, class parties, club parties	193	68	33	12
Arithmetic	192	70	43	1
Visiting stores and factories	191	85	16	14
Mystery stories like *Dracula, The Triple Murder*	190	70	35	11
Playing games like billiards and pool	188	54	36	28
Reading magazines like *Saturday Evening Post, Liberty, Colliers*	188	88	24	6
Laboratory	187	17	12	90
Riding a bicycle	187	66	24	29
Reading movie and radio magazines	185	65	31	25
Buying things to play and experiment with	185	61	24	36
Repairing mechanical things, like bicycles, autos, fuse plugs	184	46	31	45
Playing cards and other games with my parents	184	73	22	27
Reading and listening to the radio in bed	184	44	36	42
Adventure stories like *Captain Blood, Sky Birds*	183	87	27	9
Figuring what I'd get on report cards	179	77	43	7
Keeping pets like fish, pigeons, dogs, guinea pigs	177	43	42	44
Belonging to clubs like Hi-Y, DeMolay, YMCA	177	40	16	73
Arguing about political questions	177	64	50	15
Playing indoor games like chess, checkers, dominoes	177	66	58	5
Talking to my parents about my future	174	69	49	14
Assemblies	173	88	41	4
Reading columns like Neal O'Hara's and Walter Winchell's	173	67	41	25
Geography	172	96	35	3
School library	172	95	17	22
Reading newspaper editorials	171	89	38	8

TABLE VII.—*Continued*

	L, like	*I*, indifferent	*D*, dislike	*U*, unknown or not experienced
Books and articles about what engineers and surgeons do	171	79	36	20
Doing my home-work at recess and during study periods	169	85	44	8
School principal	168	106	24	8
Listening to plays and operas	168	65	68	5
English	167	97	41	1
Reading about the local and foreign political situations	167	86	47	6
Geometry	166	65	73	2
Doing jig-saw puzzles	166	72	58	10
Going to dancing parties, dinner parties, other parties	165	60	41	34
Keeping a cash account	165	46	34	61
School playgrounds	164	81	23	38
Lectures about chemical, electrical, and mechanical subjects	164	57	57	28
Student elections	163	103	25	15
Asking questions in class	161	84	61	
Doing my home-work at home	160	81	65	
Reading detective and criminal-detection magazines	159	84	41	22
Hoping that the teacher is absent	159	98	39	10
General science	158	47	22	79
Visiting other schools, junior, high, and art	158	52	17	79
French	155	67	63	21
Doing problems in class	155	100	51	
Discussing report cards with the gang	155	81	57	13
Listening in on short wave broadcasts	155	50	18	83
Helping parents at home	153	124	14	15
Listening to phongraph records	153	105	40	8
Reading the magazine and book-review sections	153	94	45	14
Reading columns like Brisbane's and summaries of news columns	150	71	38	47
Wishing there was no school	150	87	57	12
Doing school work alone	149	96	55	6
Being an officer like president or secretary of a club	149	80	39	38
Home room	148	119	31	8

TABLE VII.—*Continued*

	L, like	*I*, indifferent	*D*, dislike	*U*, unknown or not experienced
Chemistry	146	25	31	104
Visiting other clubs	146	63	29	68
Knowing the name of any auto on the road	144	104	25	23
Reading illustrated magazines like *National Geographic, Graphic*	144	59	85	18
Spelling	143	107	55	1
Auditorium	143	100	32	31
Working in a store	143	66	28	69
Visiting the fire station	143	114	20	29
Visiting girls, and going out with girls	143	82	33	48
Discussing many subjects with girls	143	81	46	36
Dreaming and musing about the problems of life	142	91	33	40
Checking my weight regularly	141	107	32	26
Western stories like *Nevada, Robbers' Roost*	141	91	57	17
Giving my opinions about teachers	141	86	61	18
Working on a farm	140	31	29	106
Histories of chemistry, science, airplanes	140	90	57	19
Watching men like carpenters and painters at work	139	111	49	7
Browsing in a book store	139	81	48	38
Reading papers in other languages like French and Italian	139	34	68	65
Books and articles about times and people, psychology	138	86	51	31
Discussing new ideas for fraternity and club	137	79	35	55
Playing a musical instrument	135	42	36	93
Repairing wooden things like furniture	135	90	36	45
Taking and developing snap-shots	135	55	28	88
Present-day novels like *Main Street, American Tragedy*	135	90	50	31
Listening to classical music and symphonies	135	81	83	7
Old-time novels like *Vanity Fair, Silas Marner, Ivanhoe*	134	97	69	6
Reading magazines like *Ballyhoo, Bunk, Snappy Stories*	134	94	57	21
Making electric and mechanical things, like spark coils	133	33	33	107
Wearing special kinds of shirts and neckties	132	109	55	10
Love-story movies like "Farewell to Arms" and "Grand Hotel"	132	75	81	18

TABLE VII.—*Continued*

	L, like	*I*, indifferent	*D*, dislike	*U*, unknown or not experienced
Helping others do their home-work	131	111	48	16
Going to church and prayer meetings	130	121	44	11
Making wooden things like cabinets, dog houses, bird houses	128	64	53	61
Going with father to his office and club	128	48	28	102
Making an especial study for an examination	128	72	95	11
Working at jobs like caddying and ushering	126	46	30	104
Listening to sales people selling things	126	66	111	3
Visiting court	125	31	14	136
Reading theatrical magazines like *Foot-lights* and *Spot-lights*	125	71	49	61
Collecting things like stamps, minerals, library cards	122	70	64	50
Discussing home-work and school affairs	122	108	64	12
Doing cross-word puzzles	122	75	90	19
Reading short articles magazines like *Forum*, *Nation*, *Golden Book*	122	95	65	24
Reciting in class, and offering to recite in class	121	113	66	6
Books and articles about foreign countries	120	111	58	17
Taking part in debates	120	53	89	44
Reading a foreign language	120	62	99	25
Watching others dance	120	112	53	21
Belonging to a science club	119	42	32	113
Group school work	118	101	50	37
Making model ships and airplanes	118	70	43	75
Listening to concerts like Damrosch's on the radio	117	86	76	27
Drawing	115	66	79	46
Driving members of the family around in the car	115	32	31	128
Accepting a dare	115	113	58	20
Singing	114	74	75	43
Belonging to journalist, debating, dramatic clubs	114	48	46	98
Belonging to Boys Scouts and Cadets	114	62	53	77
Working for the team as manager or assistant to the coach	114	59	50	83
Drawing pictures and cartoons	113	75	58	60
Entertaining people with magic and card tricks	113	70	25	98
Walking and going places alone	113	71	112	10
Shopping and window-shopping	113	74	84	35
Student government	112	88	25	81

TABLE VII.—*Continued*

	L, like	*I*, indifferent	*D*, dislike	*U*, unknown or not experienced
Kissing and holding hands with girls	112	82	48	64
Practicing a code	112	51	60	83
Listening to lectures about foreign countries	112	77	87	30
Discussing report card with my parents	111	69	120	6
Playing solitaire	109	98	75	24
Historical stories like *Oliver Cromwell, Caesar and Cleopatra*	108	96	73	29
Writing-up sporting events	107	55	62	82
Latin	106	97	98	5
Shop-work	106	63	49	88
Biographies of men like Calvin Coolidge, Al Smith, John Milton	106	128	64	8
Playing in a band or orchestra	105	28	33	140
Writing essays and themes	104	67	127	8
Working in class plays or operas	103	67	81	55
Listening to choir and organ music	103	91	105	7
Writing stories	102	72	93	39
Going to church parties	101	89	60	56
Teaching a class of younger boys	101	44	66	95
Being an officer in the student government	101	58	57	90
Keeping a regular schedule of what I am going to do	100	82	53	71
Looking up references that the teacher has given	100	98	104	4
Listening to lectures about people, like psychology lectures	98	56	113	39
School buildings	96	135	67	8
Discussing women and night life	94	94	65	53
Textbooks	93	135	71	7
Reading books about chemistry and electricity	93	66	118	29
Working for the student government	93	78	58	77
Watching shopping crowds	92	122	82	10
Reading poetry like "The Idylls of the King," "The Odyssey," "The Lay of the Last Minstrel," Kipling's poems	91	88	113	14
Listening to speeches	91	119	96	
Raising flowers and plants	90	91	69	56
Arguing with my teachers	90	74	117	25
Going to Sunday School	89	99	82	36
Listening to special hours like the Jewish and Catholic hours	89	86	98	33

TABLE VII.—*Continued*

	L, like	*I*, indifferent	*D*, dislike	*U*, unknown or not experienced
Giving driving lessons	87	27	39	153
Writing for the school paper	87	71	58	90
Sneaking into places where admission is charged	87	66	85	68
Chasing after fire-engines	85	97	77	47
Mechanical drawing	84	34	29	159
Women teachers	83	175	48	0
Betting on elections and sporting events	83	62	90	71
Histories of foreign countries	82	100	110	14
Staying for the second show	82	100	112	12
Watching girls and women	81	125	69	31
Driving an auto fastest on the road	80	73	62	91
Wondering about the teacher in class	80	102	69	55
Bumming rides	78	76	79	73
Thinking of other things in a class period	77	98	107	24
Taking part in the school opening exercises	76	120	79	31
Belonging to a mathematics club	74	53	57	122
Watching others playing cards	74	85	137	10
Staying after school for extra work or to finish work	73	92	134	7
Counsellors	72	109	34	91
Drawing pictures in class	71	83	93	58
Painting pictures	70	47	70	119
Shakespeare's plays and Lamb's *Tales from Shakespeare*	70	108	121	7
Belonging to a French club	69	50	71	116
Belonging to a poker club	69	59	49	129
Reading *Handbooks of American Prisons, American Cavalry*	69	90	111	36
Acting as a monitor around school	68	64	118	56
Writing book reports and outlines	67	92	141	6
Writing and making speeches	67	58	145	36
Working on a paper route	66	74	60	106
Looking at poultry	65	118	91	32
Reading books about animals, poultry, dairying	65	89	107	45
Reading the classified ads and stock market reports	64	70	142	30
Reading love-story magazine like *True Story*	62	63	155	26
Drinking beer, wine, and cocktails	60	48	86	112
Sleeping late and taking afternoon naps	58	65	148	35

TABLE VII.—*Continued*

	L, like	*I*, indifferent	*D*, dislike	*U*, unknown or not experienced
Talking sign language	58	58	56	134
Staying after class to talk to the teacher	58	99	97	52
Fighting with other boys and gangs	57	72	137	40
Fooling, talking, sleeping in class	56	91	123	36
Missing periods	56	44	146	60
Teasing, scaring, and annoying girls	55	63	119	69
Teasing other members of the family and the neighbors	55	62	116	73
Practicing music lessons	54	50	104	98
Discussing subjects such as bird life	54	66	120	66
Looking for migratory birds	54	83	67	102
Playing poker and black-jack for money	52	56	115	83
Histories of literature	51	100	134	21
Chewing gum and candy in class	51	83	121	51
Skipping school to go to a show	51	40	97	118
Going to choir and choir practice	50	50	60	146
Doing comedy dancing	49	43	68	146
Playing around the streets with the gang at night	47	54	136	69
Belonging to an English club	46	69	77	114
Religious books and papers like, *The Worker and the Bible*, Sunday-school and church papers	46	117	115	28
Painting lessons	46	32	54	174
Taking the family car without permission	45	31	46	184
Spending my allowance quickly	45	59	159	43
Glancing at my home-work and hurrying through it	45	89	149	23
Listening to church services on the radio	45	93	149	19
Reading books about palmistry and fortune-telling	44	91	129	42
Studying before breakfast	44	43	173	46
Playing gambling machines, nigger pools, and lotteries	43	30	110	123
Reading essays like "On Contentment"	42	77	91	96
Reading patriotic essays like "Burke's Speeches"	42	79	111	74
Belonging to a Book-of-the-Month-Club	42	64	51	149
Delivering things like samples, bills, posters	40	97	103	66
Playing cards and shooting dice for money	40	31	91	144
Pinching little things like apples and candy	40	61	106	99
German	39	17	20	230

TABLE VII.—*Continued*

	L, like	*I*, indifferent	*D*, dislike	*U*, unknown or not experienced
Selling things like tickets, papers, magazines	39	62	151	54
Examinations	38	143	123	2
Belonging to a Latin club	37	58	99	112
Arguing with my parents about going to school	37	52	99	118
Smoking at home	37	34	59	176
Non-educational books and books not good for me	35	118	119	34
Raising cain with and sassing teachers	35	84	142	45
Going to Bible-class and missionary club	34	94	61	117
Practicing the alphabet backwards	34	49	92	131
Reading operas like "Il Ponsoroso"	33	83	123	67
Bluffing through classes	32	85	151	38
Hebrew	31	18	11	246
Studying another subject in class	31	82	155	38
Elecution lessons	30	30	61	185
Physics	29	19	16	242
Writing poetry	28	35	183	60
Composing music	28	25	59	194
Copying my home-work from someone else	27	78	169	32
Getting tight	25	26	80	175
Borrowing and lending money	24	80	165	37
Just hanging around doing nothing in particular	22	30	211	43
Ringing fire-alarms and doing things to make cops mad	22	48	126	110
Reading housekeeping magazines like *Delineator*, *Home Companion*	22	40	199	45
Staying home from church to study	21	65	124	96
Hooking street-car rides	20	25	130	131
Smoking between periods	20	14	127	145
Being late for school	18	41	204	43
Getting scolded by teachers	17	57	206	26
Skipping periods for a smoke	17	13	123	153
Beating a friend out of some money	16	36	171	83
Walking to the school with the teacher	16	51	131	108
Fooling during an examination	12	31	208	55
Hitting, kicking, and teasing other boys	5	27	205	69
Spanish	0	0	0	306

Having sorted the items into groups according to frequency, we have again sorted them by the classifications which we have previously described to get the contents of Table VI. Study of this table reveals that there is a wide dispersion of likes among our divisions of interests. No school subjects are liked by more than seventy per cent of our group. Church and creative activities all come below the fifty per cent liked group. Trips and visits, athletics, and movies stand high in the list as does also reading. Anti-social activities are least liked.[1]

SELECTED BIBLIOGRAPHY

Brainard, P. P.: "Interest Tests in Vocational Guidance." *Vocational Guidance Magazine*, Vol. VI, No. 4, 1928, pp. 156–159.

Dunlap, J. W.: "Preferences as Indicators of Specific Academic Achievement." *Journal of Educational Psychology*, Vol. XXVI, No. 7, 1936, pp. 411–415.

Franklin, E. E.: *The Permanence of Vocational Interests of Junior-High-School Pupils.* The Johns Hopkins Studies in Education, No. 8, 1924, 63 pp.

Garretson, O. K.: *Relationship between Expressed Preferences and Curricular Abilities of Ninth-Grade Boys.* Teachers College, Columbia University Contributions to Education, No. 396, 1930, 77 pp.

Kelley, T. L.: *Educational Guidance; An Experimental Study in the Analysis and Prediction of Ability of High-School Pupils.* Teachers College, Columbia University Contributions to Education, No. 71, 1914, 116 pp.

Rothney, J. W. M.: "Evaluative Attitudes and Academic Success." *Journal of Educational Psychology*, Vol. XXVII, No. 4, 1936, pp. 292–298.

Segel, D.: "Differential Prediction of Scholastic Success." *School and Society*, Vol. XXXIX, No. 995, 1934, pp. 91–96.

Shuttleworth, T. K.: *The Measurement of Character and Environmental Factors Involved in Scholastic Success.* University of Iowa Studies in Education, No. 1, 1929.

Strong, E. K.: "A Vocational Interest Test." *The Educational Record*, Vol. VIII, No. 2, 1927, pp. 107–121.

Wyman, J. B.: "The Measurement of Interests." *Vocational Guidance Magazine*, Vol. VIII, No. 2, 1929, pp. 54–60.

---

[1] Notwithstanding our efforts to have the students feel that they were free to report anything they wished without any form of return, it appears that they have failed to report anti-social preferences. Another interpretation is, of course, that they do not have such preferences in large number.

A COMPARATIVE STUDY OF TWO TEACHING APTITUDE TESTS

N. FRANKLIN STUMP

Keuka College, Keuka Park, N. Y.

Careful scrutiny of the bibliographical materials reveals a scarcity of correlational studies concerning teaching aptitude tests. This fact is rather startling since it is commonly accepted that correlational studies of mental and educational tests produced a wider understanding and a deeper interpretation of their scores. What correlational studies have accomplished for mental and educational tests they can also achieve for aptitude tests.

The Morris Trait Index L[1] and the George Washington Teaching Aptitude[2] tests were administered to thirty-seven sophomores and sixty-one freshmen at Keuka College. The possibilities of this study emerged after the tests were scored, since the original purpose in securing results on these tests, along with those on other tests, was to assist the Department of Education in the selection of candidates who should be guided into professional training as prospective teachers.

The Morris Trait Index L consists of six sections: (1) Likes and Dislikes; (2) Resourcefulness of Insight; (3) Social Feelings; (4) Degree of Positiveness of Judgment; (5) Social Judgment; (6) Social Attitudes; (7) Total Scores: Leadership in General.

The George Washington Test is composed of five sections: (1) Judgment of Teaching Situations; (2) Reasoning and Information concerning School Problems; (3) Comprehension and Retention; (4) Observation and Recall; (5) Recognition of Mental States from Facial Expressions; (6) Total Scores: Teaching Aptitude.

It is the purpose of this paper to describe: (1) the comparative effectiveness of the total and part scores of the Morris Trait Index L and of the George Washington Teaching Aptitude test; (2) the intercorrelations between them; (3) the form and content which seem to show the greatest promise as measures of probable future teaching success; (4) which traits, determined by these tests, teachers should possess in order to be assured of qualities essential for teaching.

For convenience in description the Morris Trait Index L will be used as a criterion of probable future success in teaching. All comparisons will be made with this assumption in mind.

[1] Published by the Public School Publishing Co., Bloomington, Illinois.

[2] Published by Center for Psychological Service, Washington, D. C.

RESULTS

In Table I comparisons are made between the norms and quartiles of the tests and the medians and quartiles for the results secured in this study. Norms are available on the George Washington Teaching Aptitude test for high-school, normal-school, and teachers-college students, and for experienced teachers in city, town, and country systems.

The norm for the above mentioned groups, which most closely represents the median for the Keuka group, was chosen for comparison of the data on the George Washington test. For instance, the norm

TABLE I.—MEDIANS AND QUARTILE DEVIATIONS FOR THE SUBJECTS IN THIS STUDY COMPARED WITH THE NORMS OF THE GEORGE WASHINGTON TEACHING APTITUDE TEST AND MORRIS TRAIT INDEX L

	N	Md.	Q_1	Q_3	Q
George Washington Teaching Aptitude Test					
Norm-experienced town teachers.....	610	143.00	134.00	152.00	9.00
Keuka students.....	98	141.67	131.25	148.96	8.85
Morris Trait Index L					
Norm-Freshmen[1].....	228	104.30	95.52	112.68	8.58
Keuka students.....	98	110.13	102.46	118.60	8.07

[1] Note: The measures of central tendency and dispersion were calculated by the writer from the graph found on p. 4 of the Teacher's Handbook for Morris Trait Index L.

which most closely represented the Keuka median was for experienced teachers in a town system. For the Keuka group the median is 141.67, $Q = 8.85$; the experienced group, 143.00, $Q = 9.00$.

Medians and quartiles were not available, as standards for the Morris test, but in the section on "standards" of the teacher's handbook, which accompanies the test, a graph representing the frequency and scores made by two hundred twenty-eight freshmen at the New York State Teachers College is presented. From these data the writer calculated the median and quartiles in Table I. For the Keuka group the median is 110.13, $Q = 8.07$; the norm is 104.3, $Q = 8.58$.

Table II shows the intercorrelations between the totals and parts of the Morris Trait Index L and the George Washington Teaching Aptitude test, coefficients being calculated by the Pearson Product-Moment method.

If it is assumed that these tests are equally effective measures of teaching aptitude, a much higher coefficient of correlation than .35 between the total scores would be expected. While it must be recognized that aptitude tests, comparatively speaking, are in the pioneer stage as compared to mental tests, yet two different intelligence tests which would not give correlations higher than .35 would invalidate either one of them as a measure of mental ability.

The highest coefficient was found between Social Attitude scores, Section VI of the Morris Trait L and George Washington total scores, namely .536. Total scores of the Morris test and the George Washington total have an r of .35, ranking second. The third highest coefficient is .344 between the Social Attitude test, Section VI of the Morris Trait L, and Comprehension and Retention, Section III of the George Washington test. Ranking fourth is the correlation between Morris Trait total and Reasoning and Information concerning School Problems, Section II of the George Washington test, r being .324. The fifth

TABLE II.—COEFFICIENTS OF CORRELATION BETWEEN TOTAL AND SECTIONAL SCORES OF THE MORRIS TRAIT INDEX L AND GEORGE WASHINGTON TEACHING APTITUDE TEST. $N = 98$

	Washington total teaching aptitude	Washington Test I, judgment of teaching situations	Washington Test II, reasoning and information on school problems	Washington Test III, comprehension and retention	Washington Test IV, observation and recall	Washington Test V, recognition of mental states from facial expressions
Morris total, leadership in general	.35	.192	.324	.224	.107	.0614
Morris Sec. I, likes and dislikes..	.157	−.047	.198	.199	−.0571	.1485
Morris Sec. II, insight and resourcefulness.....	.161	.264	.164	.079	.053	−.0282
Morris Sec. III, social feelings...	.318	.206	.286	.23	.135	−.0386
Morris Sec. IV, positiveness of judgment.....	.277	.074	.196	.201	.212	−.0194
Morris Sec. V, social judgment...	−.0335	−.168	.048	−.008	.116	−.0216
Morris Sec. VI, social attitudes...	.536	.088	.209	.344	.030	.168

rank is between Social Feelings, Section III of the Morris test, and the George Washington total scores, r being .318. All the remaining coefficients are less than .30.

The separate tests which contribute most to the total scores of each of these two measures of teaching aptitude, which result in making the r between the total scores as large as it is, are Sections II and III of the George Washington test dealing with Reasoning and Information concerning School Problems, and Comprehension and Retention, and Section III and VI of the Morris Trait L dealing with Social Feelings and Social Attitudes; the coefficients between Sections II and III of the George Washington test and the total of the Morris Trait L being .324 and .224, respectively, while the coefficients between Sections III and VI of the Morris test and the George Washington total being .318 and .536, respectively.

It is rather interesting to note that Sections II and III of the George Washington test which show higher correlations with the criterion than do the three remaining sections of this test are composed of items of the true-false type.

Section V, which Morris regards as a test of Social Judgment, does not, by and large, show any relationship with the total scores and sections of the George Washington test. For instance, Section I of the George Washington test is intended to measure Judgment of Teaching Situations, and yet the correlations between this section and Section V of the Morris test, which are aiming at the measurement of almost identical traits, have a negative coefficient of .168. This section of the Morris test, dealing with Social Judgment, has four negative correlations with the six sections on the George Washington test. The two remaining coefficients are positive but low.

The Recognition of Mental States from Facial Expressions, Section V of the George Washington test, has doubtful value as a test of teaching aptitude when the Morris Trait L is used as criterion. Four of the six correlations are negative, ranging from .0386 to .0194. The three remaining coefficients are positive but low, ranging from .0614 to .168.

Observation and Recall, Section IV of the George Washington test, seems almost equally as ineffective as Section V, one coefficient showing a low negative correlation; namely, .0571. The remaining five correlations between this section and the Morris test are positive and range from .053 to .212, all of the correlations excepting one being below .20.

Judgment of Teaching Situations, Section I of the George Washington test, seems to rank third as a measure of teaching aptitude, using the total scores of the Morris test as criterion. Sections II and III appear superior to Section I. There are two negative coefficients between Section I and the Morris test; namely, −.047 and −.168. The five remaining coefficients are positive and range from .074 to .264.

Section I of the George Washington test is a multiple-choice test and it is surprising that it does not indicate greater promise as a measure of teaching aptitude. Sections IV and VI of the Morris test which are the multiple-choice type show considerably greater merit.

CONCLUSIONS

1. Too much emphasis cannot be placed upon social attitudes as necessary qualities to produce probable success of the teacher. About a decade ago the writer was superintendent of schools in a southern town. When he stressed correct attitudes as requisites for the successful teacher, it was noted that a tremendous increase in the efficiency of the school resulted, the progress of the pupils having been measured from month to month by means of objective educational tests. The writer believes that correct attitudes on the part of the student-teacher concerning professional training in general, student practice-teaching, etc., are the best indicators of probable future success. This is a hopeful situation so far as the teacher-training institution is concerned since the college and university staff can do much to create correct social attitudes on the part of the prospective teacher.

2. Social feelings seem essential for probable future success in teaching.

3. True-false tests based on Reasoning and Information concerning School Problems, and Comprehension and Retention of a passage which has been read a short time in advance of testing—Sections II and III of the George Washington test—seem the most effective sections of this test as measures of teaching aptitude.

4. The multiple-choice tests, Section I of the Washington test, consisting of items proposed to measure Judgment of Teaching Situations, Observations, and Recall based on a five-minute study of a picture, and Recognition of Mental States from photographs, seem effective in the order named.

5. If the George Washington test and the Morris Trait L are measures of teaching aptitude and the one is taken as a check on the other, Social Judgment, Section V of the Morris test, consistently shows little value in throwing light upon probable future success.

6. These teaching aptitude tests should be further analyzed with respect to the relationship of their total and sectional scores with measurable traits of personality, emotion, and attitude.

PREDICTING SIGHT-SINGING ABILITY IN TEACHER-EDUCATION[1]

CHARLES D. DEAN

Eastern Montana State Normal School, Billings, Montana

The purpose of this study was to determine what value the Seashore Tests of Musical Talent might have in predicting the success or failure of entering students in the Eastern Montana State Normal School in required courses in sight-singing and ear-training. The specific problems involved were:

(*a*) Can success in sight-singing be predicted by means of intelligence test scores?

(*b*) Can prediction be based on the scores made on the Seashore Tests of Musical Talent (taken individually or as a battery)?

(*c*) To what extent can the amount of previous musical training be used to predict what a student will do in sight-singing?

(*d*) What combination of the factors measured will give the best prediction of sight-singing success?

Sight-singing and ear-training as taught in the Eastern Montana State Normal School are a review of the elements of music: Notation, scale forms, intervals, simple tempos, rhythms, chords, and three- and four-part songs; the purpose being to prepare normal-school students to teach public-school music courses in the elementary schools. Group and individual drills in singing by sight are given. Each student before completing the course is required to demonstrate ability to sing a simple song without error.

Those students who fail to do satisfactory work in the sight-singing course are put into a special class during the following quarter where an attempt is made to discover and train any other musical talents they may have. In this special course much attention is given to reading and playing music without tone. This is accomplished with drum sticks, rhythm band practice, special work in elementary piano, and the teaching of school music with the use of phonograph records.

In this study an attempt is being made to see if it is possible to do away with the loss of time and effort involved on the part of both student and teacher by requiring all students, regardless of ability, to

[1] Based upon Dean, Charles D.: *The Seashore Tests of Musical Talent in Teacher-Education.* Unpublished Doctor's Field Study, Colorado State College of Education, Greeley, Colorado, 1936.

take the beginning sight-singing and ear-training course. The question is, would it not be better to place musically-handicapped students in the special class immediately on entering school rather than require them first to fail in the regular required course?

Two groups of students were used in the experiment. The first group was composed of one hundred fifty-one students who entered school in the fall of 1934 and the second group contained one hundred forty-three students who entered school in the fall of 1935. The same procedure was employed with both groups. The two groups were used in the experiment in order that the results obtained for the second group might act as a check and serve to validate the results obtained for the first group.

The Terman Group Mental Test and the Seashore Tests of Musical Talent were administered to all students during the first week of the fall quarters of 1934–1935 and 1935–1936. The amount of previous music training was determined by asking each student to indicate, on a blank provided for that purpose, the amount of time spent in courses in public-school music and under private instruction.

To determine the relation between intelligence and musical ability correlations between scores made by students on the Terman Mental Test and their composite score on the Seashore battery of tests were computed. The relation between intelligence and musical achievement was found by correlating mental test scores and marks obtained in sight-singing. Correlations between scores on the tests of the Seashore Battery, by individual tests and the test as a whole, and marks in sight-singing were computed to determine the relation of musical talent to musical achievement.

In this study, Table I, intelligence scores were found to be of but little value in predicting success in sight-singing. Prior musical training was of slightly greater value. Those factors having the greatest influence were the Seashore Pitch and Seashore Memory Tests. These tests correlated with the criterion to the extent of $.61 \pm .03$ and $.63 \pm .03$ for the first year. For the second year these correlations were increased to $.62 \pm .03$ and $.70 \pm .02$, respectively. The correlations between music marks and the Seashore Tests of consonance, time, intensity, and rhythm were all too low to be of any material value in predicting success in sight-singing.

From Table I it will be seen that of the independent variables—pitch, memory, and training—yielded higher correlations than did any of the other independent variables. The partial influences of each of

TABLE I.—CORRELATIONS OF THE INDEPENDENT VARIABLES WITH MUSIC MARKS[1]

Music mark	Year	Intelligence	Pitch	Intensity	Time	Consonance	Memory	Rhythm	Training
1	2	3	4	5	6	7	8	9	10
1	1934	.24 ± .05	.61 ± .03	.38 ± .03	.27 ± .05	.35 ± .05	.63 ± .03	.32 ± .05	.44 ± .044
2	1935	.25 ± .05	.62 ± .03	.40 ± .05	.22 ± .05	.29 ± .05	.70 ± .03	.39 ± .05	.43 ± .045

[1] The scores for 1934 are based on one hundred fifty-one cases and those for 1935 on one hundred forty-three cases.

the four significant variables (music mark, pitch, tonal memory, and training) were computed separately for each of the two studies; that is, for the class of 1934 and the class of 1935, and multiple cor-

TABLE II.—MULTIPLE AND PARTIAL CORRELATIONS BASED ON DATA OBTAINED FROM ONE HUNDRED FIFTY-ONE AND ONE HUNDRED FORTY-THREE ENTERING STUDENTS AT THE EASTERN MONTANA STATE NORMAL SCHOOL, BILLINGS, MONTANA, FALL QUARTERS OF 1934 AND 1935 RESPECTIVELY

Variables[1]	Coefficients	
	(1934)	(1935)
1	2	3
$R_{1(23)}$	.69 ± .029	.74 ± .025
$R_{1(234)}$	.72 ± .026	.77 ± .024
$r_{12\cdot3}$	.37 ± .05	.35 ± .05
$r_{14\cdot3}$	.33 ± .05	.27 ± .05
$r_{24\cdot3}$	.24 ± .05	.17 ± .05
$r_{13\cdot2}$	.41 ± .05	.52 ± .04
$r_{14\cdot2}$	.29 ± .05	.30 ± .05
$r_{34\cdot2}$	.12 ± .05	.21 ± .05
$r_{12\cdot34}$	.32 ± .05	.32 ± .05
$r_{13\cdot24}$	.39 ± .05	.49 ± .04
$r_{14\cdot23}$	.27 ± .05	.23 ± .05

[1] 1. Mark in sight-singing.
2. Seashore Pitch score.
3. Seashore Memory score.
4. Years of musical training.

relations calculated. These results, Table II, showing correlations of .69 and .74, respectively, for the two years between the sight-singing mark and the Seashore scores of pitch and memory are approximately

the same as those obtained by Salisbury and Smith[1] at the Bellingham State Teachers College. They found a multiple correlation of .723 between sight-singing and pitch and tonal memory scores where a similar course in sight-singing was used as the criterion for musical success. These data seem to validate the Salisbury and Smith study and present further evidence to show that prediction of success in sight-singing can be somewhat determined in advance by administering the Seashore Pitch and the Seashore Memory Tests. A further examination of the results indicates that a little is gained by taking into consideration the fourth variable, musical training; the increase in the multiple correlation being from .69 to .72 in the first study and from .74 to .77 in the second.

On the basis of the results obtained in this study we can predict the student's most probable grade or mark by using the regression equation in its score form. By using this equation, weights can be assigned to each of the independent variables in such a way that the dependent or criterion variable can be forecasted with a minimum of error. In writing these regression equations, X_1—music mark—will be the dependent or criterion variable, X_2 will represent the score on the pitch test, X_3 will be the tonal memory score, and X_4 will be used to designate the number of years of musical training. For the class of 1934–1935, involving one hundred fifty-one students, the equation becomes $X_1 = .012X_2 \pm .015X_3 \pm .05X_4 - .18$ in which $R_{1(234)}$ is equal to .72 and the PE $_{(est.\ x1)}$ is equal to $\pm .53$. The regression equation for the 1935–1936 study, involving one hundred forty-three students, is very much like that for 1934–1935. In this case the equation is $X_1 = .011X_2 \pm .019X_3 \pm .05X_4 - .44$. In this case $R_{1(234)}$ is equal to .77 and the PE $_{(est.\ x1)}$ is equal to $\pm .51$.

In all of these equations the probable error of estimate is roughly one-half of a point on a five point scale. This means that the chances are even that the most probable mark that a student will receive in sight-singing will not vary over one-half a point either way from the obtained or estimated score. We may also be practically certain that any estimated score will not be in error by more than two points.

If we assume that the independent variables caused the result expressed by the dependent variable, in this case a certain mark in music, we might express the results in terms of the percentage of con-

[1] Salisbury, F. S., and Smith, H. B.: "Prognosis of Sight-singing Ability of Normal School Students." *Journal of Applied Psychology*, Vol. XII, 1929, pp. 425–439.

tribution of the former to the latter by squaring the coefficients of the regression equations in the "special form" with sigmas equal to one, or by squaring the path coefficients obtained. By using path coefficients the influence of the independent variables was found to be substantially the same as was found by using the partial correlation technic.

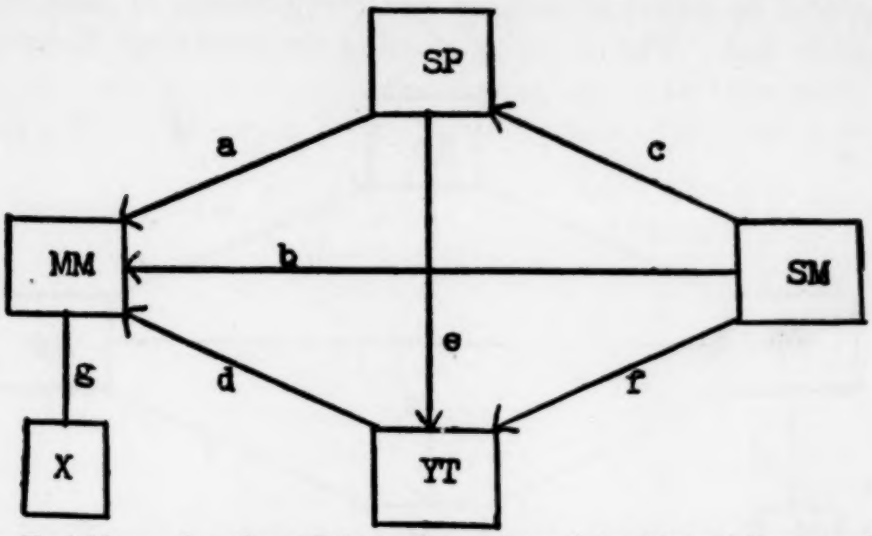


FIG. 1.—Variables and path coefficients drawn from data obtained from one hundred fifty-one Eastern Montana State Normal School students, 1934–1935.

In this diagram:

MM = music mark received in sight-singing.
SP = Seashore Pitch score.
SM = Seashore Memory score.
YT = years of musical training.
X = all factors not included in *SP*, *SM*, and *YT* that influence the music mark.

The lines in this diagram represent the path coefficients. The paths are labeled with small letters. What they represent is shown by the following equations:

$$P_{MM \cdot SP} = a \qquad P_{SM \cdot SP} = c$$
$$P_{MM \cdot SM} = b \qquad P_{SM \cdot YT} = f$$
$$P_{MM \cdot YT} = d \qquad P_{SP \cdot YT} = e$$
$$P_{MM \cdot X} = g$$

The value of the path coefficients was found to be as follows:

1. For the 1934–1935 group, Fig. 1, four variables (music mark, Seashore Pitch, Seashore Memory, and years of training)

$$a = .3020$$
$$b = .3830$$
$$d = .2090$$

2. For the 1935–1936 group, Fig. 2, four variables (music mark, Seashore Pitch, Seashore Memory, and years of training)

$$a = .2791$$
$$b = .4713$$
$$d = .1676$$

By using this technic, however, the percentage of contribution found will be only that of the influence of the direct paths upon the independent or criterion variable and the influence of joint relationship will be lost. The necessity of using the coefficient of determina-

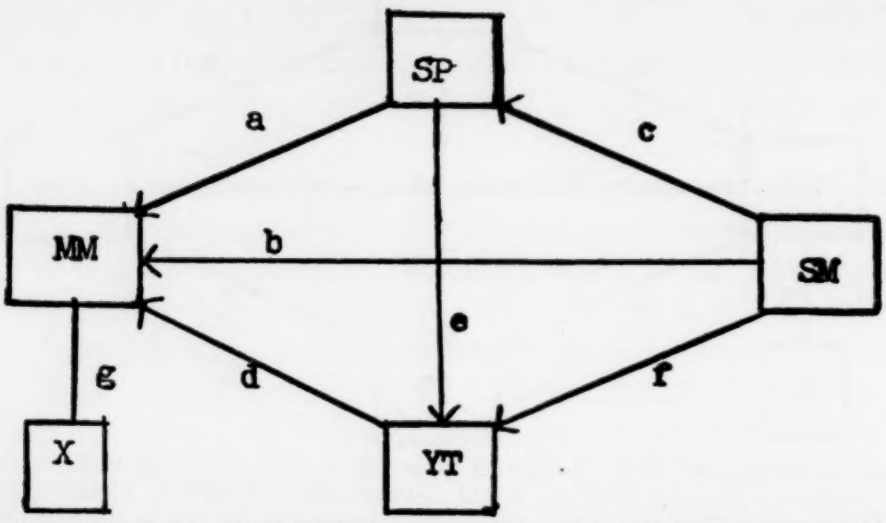


FIG. 2.—Variables and path coefficients drawn from data obtained from one hundred forty-three Eastern Montana State Normal School students, 1935–1936.

In this diagram:

MM = music mark received in sight-singing.
SP = Seashore Pitch score.
SM = Seashore Memory score.
YT = years of musical training.
X = all factors not included in *SP*, *SM*, and *YT* that influence the music mark.

The lines in this diagram represent the path coefficients. The paths are labeled with small letters. What they represent is shown by the following equations:

$$P_{MM \cdot SP} = a \qquad P_{SM \cdot SP} = c$$
$$P_{MM \cdot SM} = b \qquad P_{SM \cdot YT} = f$$
$$P_{MM \cdot YT} = d \qquad P_{SP \cdot YT} = e$$
$$P_{MM \cdot X} = g$$

tion to show the total influence of the causal factors upon the criterion or independent variable is set forth by Wright.[1] He says, "It can be shown that the squares of the path coefficients measure the degree of determination by each cause. If the causes are independent of each other, the sum of the squared path coefficients is unity. If the causes are correlated, terms representing joint determination must be recognized. . . . The squared path coefficients and the expressions for joint

[1] Wright, Sewell: "Correlation and Causation." *Journal of Agricultural Research*, Vol. XX, July, 1921, pp. 557–585.

determination measure the portion of the squared standard deviation of the effect due to the causes singly and conjointly, respectively."

The equations for the coefficients of determination are as follows:

$$d_{X \cdot A} = P^2_{X \cdot A}$$
$$d_{X \cdot \overline{AB}} = 2P_{X \cdot A}P_{X \cdot B}r_{AB}$$
$$\Sigma d_{X \cdot A} + \Sigma d_{X \cdot AB} = 1$$

In the case of our four variable problems involving music mark, Seashore Pitch, Seashore Memory, and years of musical training we find:

1. For the 1934–1935 group

$$d_{MM \cdot SP} = a^2 = .0912$$
$$d_{MM \cdot SM} = b^2 = .1467$$
$$d_{MM \cdot YT} = d^2 = .0437$$
$$d_{MM \cdot \overline{SPSM}} = 2_{ab}r_{SPSM} = .1388$$
$$d_{MM \cdot \overline{SPYT}} = 2_{ad}r_{SPYT} = .0467$$
$$d_{MM \cdot \overline{SMYT}} = 2_{bd}r_{SMYT} = .0496$$
$$d_{1 \cdot X} = g^2 = \frac{.4833}{1.0000}$$

2. For the 1935–1936 group

$$d_{MM \cdot SP} = a^2 = .0778$$
$$d_{MM \cdot SM} = b^2 = .2221$$
$$d_{MM \cdot YT} = d^2 = .0281$$
$$d_{MM \cdot \overline{SPSM}} = 2_{ab}r_{SPSM} = .1576$$
$$d_{MM \cdot \overline{SPYT}} = 2_{ad}r_{SPYT} = .0318$$
$$d_{MM \cdot \overline{SMYT}} = 2_{bd}r_{SMYT} = .0537$$
$$d_{1 \cdot X} = g^2 = \frac{.4289}{1.0000}$$

If we would distribute to the direct paths the joint influences in proportion to the values of the direct paths, we would find the total influence of pitch on music mark for the class of 1934–1935 to be 17.59 per cent, the influence of memory 27.06 per cent, and the influence of years of musical training 7.02 per cent. For the class of 1935–1936 these per cents of influence on music mark would be 14.22, 38.64, and 4.25, respectively. Thus the total influence of the Seashore Tests on the mark in sight-singing for the two years was about forty-five per cent and fifty-two per cent, respectively. These results compare favorably

with the fifty per cent distribution of the Seashore Tests to sight-singing mark found by Chadwick in his study of thirty-nine music majors.[1]

SUMMARY

The intelligence of the student is not a very important element in predicting his success in courses in sight-singing. Prior musical training is of slightly greater value. Seashore Tests of pitch and memory were found to be the most significant for predictive purposes. The scores made on these two Seashore Tests accounted for from forty-five to fifty-two per cent of the student's mark in music. Previous musical training accounted for from seven to four per cent. When pitch, memory, and years of musical training were combined from fifty-two to fifty-seven per cent of the mark made in sight-singing by Eastern Montana State Normal School students was accounted for. This leaves only about forty-three to forty-eight per cent to be accounted for by other unmeasured factors. When these measures are combined by use of the regression equation in the score form they can be used to predict the student's most possible score in sight-singing. If this is done, the chances will be even that the predicted score will not vary by more than one-half a grade point on a five point scale from the true score and in no case will the true score vary from the estimated score by more than two points. On this basis it would seem advisable to place all students whose predicted scores fall two PE or more below the median score in special classes rather than require them to take the conventional courses in sight-singing.

---

[1] Chadwick, J. Elbert: "Predicting Success in Sight-Singing." *Journal of Applied Psychology*, Vol. XVII, Dec., 1933, pp. 671–674.

HIERARCHICAL SKILLS IN TYPEWRITING

PAUL FENDRICK

The State College of Washington

This study[1] presents an attempt to measure the hierarchical relationship of certain common typewriting skills which are manifested by operators representing various levels of achievement. Its origin can be traced to the classic investigations of Bryan and Harter[2] in the field of telegraphy and Book[3] in the field of typewriting. Such terms as letter, word, and phrase habits are landmarks in the language equipment of the educational psychologist and it was, therefore, considered desirable to revaluate the hierarchical point of view by experimental typewriting under conditions somewhat at variance with the usual test procedures.

Test copy was utilized that ranged in complexity from discontinuous rare letters to connected discourse made up from the five hundred commonest words. The six intervening tests were constructed from material that could provide avenues for evaluating various aspects of typing development. These eight tests were given under uniform procedures to six groups of typists representing a wide range of typewriting proficiency.

The specific questions in this inquiry were: (1) Does the expert typist owe his high performance to a mastery of all letters with a relatively equal degree of proficiency? (2) How marked is the difference in performance of an expert when attention is directed from the letter to the word performance? (3) Are groups of typists at different levels of ability affected very much in their performance when sense copy is broken up into material that contains the same word elements? (4) What difference, if any, lies in typing copy exclusively containing words of highly common usage when related to copy containing words of less common occurrence? (5) In general, how significantly does a group of typists representing a group of extreme ability

---

[1] The report is a condensation of a thesis on file at the Yale University Library. The writer is indebted to the late J. C. Chapman both for the problem and advice on technical procedures. Professor Chapman was preparing the study for publication under a joint authorship when interrupted by his untimely death.

[2] Bryan, W. L. and Harter: "Studies in the Physiology and Psychology of the Telegraph Language." *Psychological Review*, Vol. VI, July, 1899, pp. 345–375.

[3] Book, W. F.: *The Psychology of Skill.* New York: Gregg Publishing Company, 1925, pp. viii and 257.

differ from other groups in copying easy connected discourse, word jumbles, and letter jumbles?

TEST COMPOSITION

The test battery was made up of three tests of connected discourse, two tests of word jumbles, and three tests of letter jumbles. Two of the three tests involving ordinary copy were abstracted from a random text[1] and could well be considered material that constituted copy of average difficulty. The principal selective criteria was to choose material that was free from an excess of punctuation. In this way technical forms with intricate and special spacings were avoided. These two connected discourse tests (CMPA and CMP) were considered to be on the same level of difficulty and served to check the consistency of performance of ordinary copy material. The third test involving connected discourse (CEP) was copy made up from the five hundred most frequently used words[2] in the English language.

Copy containing two kinds of word jumbles was obtained by scrambling Tests CMP and CEP, *i.e.*, one of the ordinary discourse passages and the test copy made up from the five hundred commonest words. Any marked influence inherent in copy containing only highly common words could then be determined for various levels of typing proficiency. In both instances, however, connected discourse was destroyed and only word content was to be transcribed.

Three passages containing letter jumbles were introduced to measure relative performance on the letter habit level. Test LEP consisted in the random rearrangement of the various letters in Test CEP (continuous discourse composed from highly frequent words). The other letter jumble test passages were based on the CEP pattern but differed with respect to the letter elements. Hoke[3] has reported the relative frequencies in usage of the various letters in the alphabet; consequently, Test LF was constructed utilizing only the frequent letters *ETAOSINRHL* and Test LR was derived from the least frequent letters *PWFGVKJXQZ*. The frequency of the letter elements

[1] Chapman, J. C. and Counts, G. S.: *Principles of Education.* Boston: Houghton Mifflin Company, 1924, pp. xviii and 645. (Passages began on pages 314 and 320.)

[2] Thorndike, E. L.: *The Teacher's Word Book of* 10,000 *Words.* New York: Teachers College, Columbia University, 1921, pp. 134.

[3] Hoke, R. E.: *The Improvement of Speed and Accuracy in Typewriting.* Johns Hopkins University Studies in Education, No. 7. Baltimore: Johns Hopkins Press, 1922, p. 118.

in these latter tests was pure chance, since letter sequence was determined by lot.

These tests were multigraphed uniformly in lines averaging about seventy strokes each. Single spacing was used in order to avoid the danger involving in turning the copy. This arrangement is justified because typists very often copy from this kind of manuscript.

TESTING

In selecting the subjects for this experiment the first criterion was to utilize only those individuals employing the touch method. As the touch method is uniformly taught in schools of commerce, a cross section of subjects possessing the ordinary levels of ability was available in the New Haven (Conn.) Commercial High School. There were no individuals in this population in the high-school group possessing expert ability in the sense that an expert is a person who does demonstration typing. Through the coöperation of the Underwood Typewriter Company it was possible, however, to secure eight subjects in the expert class. These typists[1] were very close to the physiological limit as far as the typing function was concerned.

Furthermore, because the available population at the New Haven High School was too large, only ninety-seven subjects, selected on the basis of the Blackstone Stenographic Proficiency Test, were studied. This method was used merely to procure a representative sampling of various levels of typing ability and did not directly determine the basis on which the classification was made for purposes of this study. Starting from a high-school group which would be classified almost as beginners, the non-expert typing students then represented various gradations of skill and culminated in the highest stage of typewriting ability commonly found after one and one-half years of training.

In giving these tests the participants were first asked to insert paper in their machines and practice for a short time on any available copy material. This procedure served as a check on the working order of the machines as well as providing a warming up period. After a few moments they were given the word to stop and the following directions:

[1] The names of these expert typists are: Ruth Martin, Rosalie Freda, Josephine Pitisan, George W. Gaskill, Barney Stapert, Arthur Neuenhaus, George Hossfield, and Bessie Friedman. At this time Mr. Hossfield's record was one hundred forty-six words per minute and Miss Friedman's performance was one hundred forty-three words per minute.

You are about to be given a series of tests. Some of this copy is made up of straightforward material and some of it you may find in code. Do your best, remembering that errors count against you. See to it that your machines are set at about seventy spaces to the line. . . . Each test will be five minutes in length. I will count three and say *go*. Stop immediately when I say *stop*.

The tests were given in the following order:

Test	Name	Beginning
1	CMPA	"Our great industrial and urban . . . "
2	WEP	"Us talk wanted to with . . . "
3	CEP	"You will remember our early . . . "
4	LEP	"Uyo llwi emrebmer uro . . . "
5	CMP	"Since literature has long been . . . "
6	WMP	"Been important an subject long . . . "
7	LF	"Lon raeh tsinaile tro . . . "
8	LR	"Pjg fxzw qkvgxupz tjq . . . "
9	CMPA	(repeated for reliability)

Test CMPA was repeated so that greater reliability might be obtained. The average performance on this test was used to secure ability levels in typewriting. This test copy, as previously stated, was of moderate difficulty and its context might well duplicate the requirements of ordinary work conditions.

CLASSIFICATION AND RESULTS

Levels of attainment were obtained by averaging the two scores made on the CMPA classification test. The resulting distribution was then divided into six categories and constituted the various levels of typing proficiency. These categories determined the basis for subsequent analytical considerations regarding the several typing skills. Table I presents this classificatory distribution and indicates the levels of typing performance for the one hundred five cases used in this study.

It is now possible to consider the performance of each of these groups of typing attainment on copy requiring a variety of correlated skills. Table II presents the means and standard deviations of the several classification groups in terms of strokes on gross[1] performance

[1] An empirical method of scoring was utilized in order to achieve the most economical method of grading the mass of accumulated data. Since the point

TABLE I.—LEVELS OF TYPING PERFORMANCE ON CLASSIFICATION TEST

Strokes	Cases	
2900	XX	
2800	X	
2700	X	
2600	X	
2500	X	Group VI (N = 8)
2400		
2300	X	
2200		
2100	X	
2000		
1900		
1800		
1700		
1600		Group V (N = 20)
1500	X	
1400	X	
1300	XXX	
1200	XXXXXX	
1100	XXXXXXXXX	
1000	XXXXXXXXXXXX	Group IV (N = 24)
900	XXXXXXXXXX	
800	XXXXXXXXXX	Group III (N = 18)
700	XXXXXX	
600	XX	
500	XXX	(Group II (N = 16)
400	XXXXXXXXXXXX	
300	XXXXXXXXXXX	Group I (N = 19)
200	XXXXXXX	

in the several experimental tests. This table forms the basis from which the subsequent analyses of the study will be made.

attained in total achievement is not a satisfactory index, an estimate of the total number of strokes was obtained by means of a specially improvised scale that measured the total number of strokes in each line. This method on several experimental trials proved to be sufficiently accurate for purposes of this study. Furthermore, a supplementary study on errors did not give any indication of significantly altering these gross scores.

TABLE II.—TYPING SKILL SCORES ON CONNECTED DISCOURSE, WORD JUMBLES, AND LETTER JUMBLES OBTAINED FROM GROUPS REPRESENTING VARIOUS LEVELS OF TYPEWRITING ABILITY

Test.....	CMPA		CMP		CEP		WEP		WMP		LEP		LF		LR	
Strokes.....	*M*	σ	*M*	σ	*M*	σ	*M*	σ	*M*	σ	*M*	σ	*M*	σ	*M*	σ
Group I.....	332	41	332	40	354	50	310	49	340	45	232	37	251	34	203	34
Group II....	466	57	459	50	482	54	435	56	448	62	336	36	350	44	258	44
Group III...	796	70	795	84	820	93	750	66	763	78	557	82	586	64	368	38
Group IV....	996	55	970	58	1041	81	953	84	948	70	695	98	747	66	456	46
Group V.....	1262	117	1220	123	1300	132	1226	132	1208	141	829	63	907	93	544	63
Group VI....	2665	[1]	2660	...	2815	...	2725	...	2585	...	1485	..	1555	..	830	

[1] Inasmuch as Group VI had only eight cases and represented typing near the physiological limit of performance, no great significance is obtained from any measure of scatter.

TYPING LETTERS BASED ON FREQUENCY USAGE

At the outset an analysis might well be made of the lowest habit order, *i.e.*, letter habits, so as to ascertain what contributions the reading and typing of the various letters make to the attainment of expertness in the function as a whole. Is the expert typist just as efficient in typing rare letters as frequent letters? Do beginners vary more than experts in typing letters of varying degrees of frequency?

Although percentage analyses are dangerous, a relationship could reasonably be established between the several classificatory groups and performance on the letter jumble tests. By comparing the percentage of superiority of the expert group on Test LF (frequent letters) in terms of his own performance on Test LR (rare letters) a gain of eighty-seven per cent is noted. On the same comparative analysis the elementary group (Group I) shows a difference of only twenty-four per cent. The formula employed to obtain these ratios was: $\frac{\text{Strokes LF} - \text{Strokes LR}}{\text{Strokes LR}}$. Table III indicates the relative differences in performance for all groups.

The evidence in Table III indicates that the expert group (Group VI) does not achieve its position through an equal mastery of reading and typing all letters, but to a relative superiority on easy letters as determined by frequency of usage. The expert has practiced the frequently occurring letters so much more that his ability must undoubtedly be affected by this difference in practice.

TABLE III.—DIFFERENCES IN TYPEWRITING RARE AND FREQUENT LETTERS

Group	I	II	III	IV	V	VI
$\frac{LF^* - LR}{LR} \times 100$, per cent	24	36	60	64	67	87

* The test code in this and all subsequent formulae should be understood to represent mean typing strokes.

Another approach to this consideration lies in contrasting the performance of Groups VI and I on Tests CMP and LR. From Table II the evidence is that the expert group (VI) drops sixty-nine per cent in typing rare letters while the elementary group (I) drops only thirty-nine per cent in the same comparison. If, however, the scores on tests CMP and LF are compared, the expert group drops forty-one per cent in typing frequent letters while the elementary group drops twenty-four per cent. The expert group, consequently differs twenty-eight per cent in its absolute decline in typing letters ranked by frequency of appearance in contrast with a decline of fifteen per cent for the elementary group. This result tends to substantiate the findings in the preceding analysis.

TYPING HIGHLY FREQUENT WORDS AND DERIVED LETTER JUMBLES

Higher orders of habits are destroyed when typing copy configurations are derived from jumbled letter constituents of common words. The facilitation due to the organization of letters into word habits might be obtained by comparing the achievement on Tests WEP (frequent words) and LEP (Letter jumbles based on CEP). Table IV based on the formula, $\frac{\text{Strokes WEP} - \text{Strokes LEP}}{\text{Strokes LEP}}$ indicates the relative changes that ensue when typing attention is redirected from word to letter copy.

TABLE IV.—DIFFERENCES IN TYPING HIGH FREQUENCY WORDS AND DERIVED LETTER JUMBLES

Group	I	II	III	IV	V	VI
$\frac{WEP - LEP}{LEP} \times 100$, per cent	34	30	35	37	48	83

Inasmuch as the most frequent letters undoubtedly predominate in Test LEP (based on five hundred commonest words), the evidence from Table IV indicates that advancement in typewriting not only proceeds with greater efficiency on the more frequently occurring letters, but also at the instance of frequent letter combinations. The advantage derived from frequent letter patterns—words—increases the performance of the expert group by eighty-three per cent while the beginners' group (I) shows a gain of only thirty-four per cent. Performance of groups in the lower grades of typing ability is noticeably less influenced by typing copy letter jumbles derived from highly common words.

TYPING CONNECTED AND UNCONNECTED DISCOURSE

If the mean achievement on a continuous discourse is compared with disconnected words of the same content, *i.e.*, sense continuity having been destroyed, some measure of the value of phrase or larger thought habits should be isolated. Table V-*a* indicates the findings related to Tests CEP and WEP.

TABLE V-*a*.—DIFFERENCES IN TYPING CONNECTED AND DISCONNECTED DISCOURSE (Commonest Words)

Group	I	II	III	IV	V	VI
$\frac{CEP - WEP}{WEP} \times 100$, per cent	14	11	9	9	6	3

The evidence in Table V-*a* makes it difficult to justify the position that the expert obtains any great superiority over the novice from connected discourse. The expert group has evidently reached a point very near the physiological limit on connected discourse copy and only a slight retardation in attainment is manifest when the copy had been reduced to word jumbles. It is true that copy containing connected discourse did manifest some advantage, but the gain was slight for the

TABLE V-*b*.—DIFFERENCES IN TYPING CONNECTED AND DISCONNECTED DISCOURSE (Unselected Words)

Group	I	II	III	IV	V	VI
$\frac{CMP - WMP}{WMP} \times 100$, per cent	−2	+2	+4	+2	+1	+3

expert group. It is quite possible, however, that the percentage differences might well involve a fluctuating significance at different levels of ability. Another approach in this consideration can be made by making similar comparisons on Tests CMP and WMP as presented in Table V-*b*.

It is evident from the data in Table V-*b* that relatively the higher levels of typewriting performance are not impeded to any great degree when word jumbles are being typed in lieu of regular copy. While sense continuity appears to facilitate typing performance very slightly, it is quite unlikely that the expert typists receive from it highly significant hierarchical advantage.

TYPING MOST COMMON WORDS AND UNSELECTED WORDS

Since Test CEP contained only the five hundred commonest words in the vernacular and Test CMP involved less common words, a comparison of these two tests should yield information regarding the influence of frequency of word occurrence on typing proficiency. It should be remembered, however, that Test CMP also contained many common words. Table VI-*a* contains these data.

TABLE VI-*a*.—DIFFERENCES IN TYPING COMMON WORDS—CONNECTED DISCOURSE

Group	I	II	III	IV	V	VI
$\frac{CEP - CMP}{CMP} \times 100$, per cent	7	5	3	7	7	6

This analysis indicates that the relative percentage of advantage in typing very common word copy persists at all levels of typewriting ability. The figures tend to negate the possibility of ascribing to the skilled typist any advantage due to greater practice on the most common words. On a priori grounds the expert might be expected to encounter the five hundred commonest words much more frequently and thereby gain considerable superiority through extended learning. This condition would be rather difficult to reconcile were it not for the possibility of making another approach to this problem. Table VI-*b*, consequently, attacks the same condition on the basis of typing disconnected word copy.

These data strongly suggest that higher levels of typing achievement are benefitted by the presence of common word habits in the

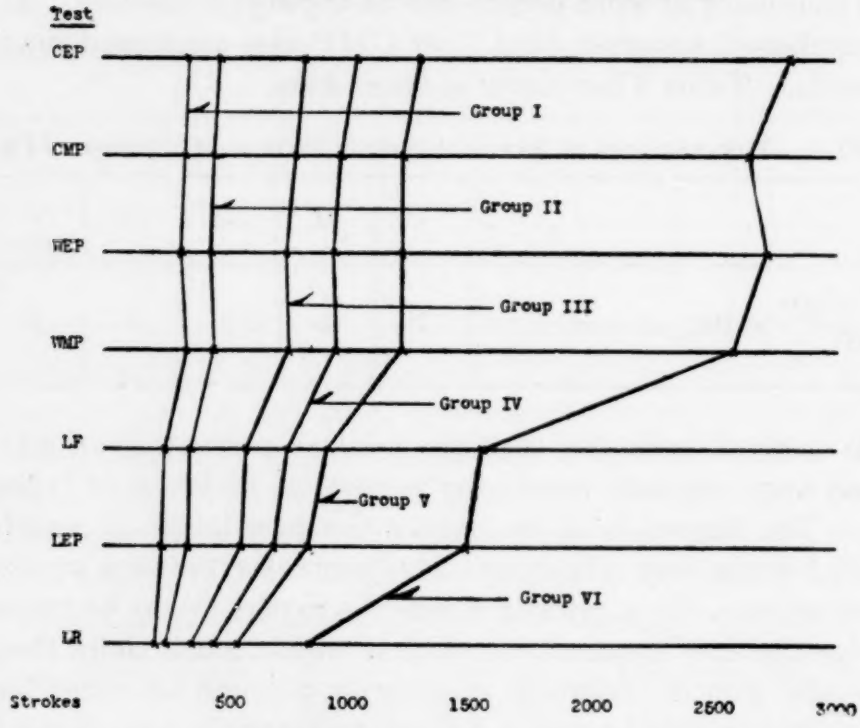
TABLE VI-*b*.—DIFFERENCES IN TYPING COMMON WORDS—DISCONNECTED DISCOURSE

Group	I	II	III	IV	V	VI
$\frac{WEP - WMP}{WMP} \times 100$, per cent	−9	−3	−2	+1	+2	+5

copy. The previous inconclusive evidence (in Table VI-*a*) might well indicate that connected discourse copy had masked the evidence relating to the mastery of highly common words. Furthermore, the consistent superiority of performance by the expert typists (Group VI) in both analyses strongly indicates a likelihood of more rigid consolidation of common word habits at this level. Certainly the oscillations for the

Figure 1

Typing Proficiency at Various Skill Levels on Letter, Word and Continuous Discourse Copy



less skilled typists are plausible in light of our knowledge of learning. As contributory evidence, the influence of common word connected discourse evidenced in Table V-*a* might serve to give a relative advantage to the less able typists and thereby compensate for the discrepancies in Tables VI-*a* and VI-*b*.

GENERAL IMPROVEMENT IN TYPEWRITING RELATED TO ALL LOWER ORDERS OF HABITS

A significant question is whether general improvement in typewriting ability involves a continuous improvement in all the lower orders of habits. Figure 1 portrays graphically the answer to this query and indicates that there is a continuous improvement of abilities on the different skills as represented by the experimental tests. This is confirmatory to our knowledge of learning because it demonstrates that progress is never limited at any particular stage. It is now possible to proceed to a more critical point.

The leading question is whether growth is proportional in each of the lower habit orders. The possibilities for this discussion will be confined for the sake of brevity only to the growth of the two extreme ability groupings, *i.e.*, Groups I and VI. Table VII will illustrate this growth. The final column preceding it indicates the gross change.

TABLE VII.—GROSS IMPROVEMENT IN TYPEWRITING SKILLS—GROUP VI COMPARED WITH GROUP I

Test	Group VI score / Group I score	Relative change	Gross change	Relative percentage change
CMP	$\frac{2660}{332}$	$\frac{2660 - 332}{332}$	8X	701
CEP	$\frac{2815}{354}$	$\frac{2815 - 354}{354}$	8X	695
WEP	$\frac{2725}{310}$	$\frac{2725 - 310}{310}$	8.8X	779
WMP	$\frac{2585}{340}$	$\frac{2585 - 340}{340}$	7.6X	660
LEP	$\frac{1485}{232}$	$\frac{1485 - 232}{232}$	6.4X	540
LF	$\frac{1555}{251}$	$\frac{1555 - 251}{251}$	6.2X	520
LR	$\frac{830}{203}$	$\frac{830 - 203}{203}$	4.1X	310

An examination of these comparisons shows that typing superiority is not due to any constant proportional growth of all the habit orders. High attainment appears to be a function of the learning and consolidation of habits involving frequent letter combinations and common words.

CONCLUSIONS

Typing copy involving continuous discourse, letter and word jumbles was given in special copy form to typists representing various levels of proficiency. The results were treated in various ways and the following conclusions are tentatively established:

(1) Increased achievement in general typewriting ability is paralleled by a steady increase in all lower order habits.

(2) These increases are, however, not proportionate for all levels of ability.

(3) Meaning derived from copy containing connected discourse does not appear to have much significance in relationship to high levels of attainment.

(4) Experts do not derive their superiority through a relative increase in speed on rare letters.

(5) Experts are greatly facilitated in their performance with copy containing frequent letter combinations or common words.

A STUDY OF THE INFLUENCE OF FAMILIARITY WITH ENGLISH WORDS UPON THE LEARNING OF THEIR FOREIGN LANGUAGE EQUIVALENTS

FLORENCE LUCILE CHAPMAN

Graduate Student, University of California

AND

LUTHER C. GILBERT

University of California

It is the purpose of this article to report a study undertaken to discover whether a foreign language vocabulary is learned more readily when the English equivalents of the foreign language words are known or when they are unknown. The criteria included growth, amount and retention of learning for a selected foreign language vocabulary. So far as can be determined, no study covering this point has been reported previously.

Subjects of the experiment were one hundred twenty-one students of at least junior standing in the field of education at the University of California. The subjects were divided into two groups, one of sixty-eight and one of fifty-three. All were of the white race and of approximately average college age. There were nine men in the larger group, twenty-one in the smaller.

To eliminate the possibility of familiarity with the critical foreign words the Hindustani language was chosen: No subject had any knowledge of that language. To control the factor of word length, Hindustani words were paired arbitrarily with English words. The establishing of false relationships appeared inconsequential in view of the slight probability that any subject would undertake to master Hindustani. As experimental technique the procedure appeared justifiable because no subject was aware of the artificiality of the relationship.

In order to obtain a list of known English words, twenty-four seven-letter nouns were selected from the five thousand group in the Thorndike *Twenty Thousand Word List.* To obtain the unknown English words, a large number of unusual seven letter words was taken from an unabridged dictionary and presented to a group of graduate students for consideration: In the light of their discussion there evolved a list of twenty-four words. Subjects of this experiment were tested on their knowledge of these words before the learning took place.

Seven-letter Hindustani words were chosen at random from an English-Hindustani dictionary: Their meanings were not considered. English and Hindustani words were paired in chance order.

Presentation of the words for learning was accomplished by means of a projection machine. Each pair of words was placed on its own slide, the English word appearing first. The words were typed with black carbon on clear cellophane and were uniformly spaced on the slide.

Prior to the learning period all subjects were tested on their familiarity with the presumably unfamiliar English words. After the learning period all subjects were tested on the English-Hindustani associations they had formed. Separate tests were provided for familiar and for unfamiliar English words, and two variations of each test were prepared to check against familiarity with the form of the test material. Each test was divided into six groups of items, each group containing four English words and six possible Hindustani equivalents. All scores were in terms of errors.

As a preliminary to the actual collection of data, subjects were introduced to the technique in a trial set-up with different material two days before the actual experiment was conducted.

Division of the total of one hundred twenty-one subjects into two groups was undertaken for the purpose of establishing the validity and reliability of the findings. The procedure was identical for both groups except that the English words for the second group had their Hindustani equivalents changed so that the equivalents of the English familiar words of one group became the equivalents for the English unfamiliar words for the other group. This procedure was intended as a check against the influence of artificial connotations between the word pairs.

For the vocabulary pre-test subjects were given the following directions:

> You will be given a sheet containing twenty-four English words to which I wish to secure your response. If you know the meaning of the word write the definition in the space provided. If you feel that you are familiar with the word make some attempt to define it or show its use in a sentence. If you do not recognize any of the words leave your paper blank, but be sure to hand it in.

When all the papers were collected, the following instructions were given:

> Forty-eight English words with their Hindustani equivalents will be flashed upon the screen. Each slide will be shown for five seconds. You

are to learn the Hindustani equivalents so that you will be able to take a multiple choice test after each presentation. The forty-eight words and their equivalents will be presented three times and you will be tested after each complete running. Two types of English words will be presented. One set will be the words you have just tried to define and the other half will be English words you will at once recognize and know the meaning of. Learn as many as you can during each running.

In order to avoid the effect of practice resulting from learning one type of word before the other, the slides were arranged so that familiar and unfamiliar English words alternated. After each presentation the slides were arranged in chance order to prevent the association of subsequent pairs and hence the anticipation of pairs to come.

Immediately following each of the three presentations a test for all words was administered. The three test forms were modified so that no two were identical. A fourth form of the test was given on the fourteenth day after the presentations.

Exactly the same procedures were used with Group One and Group Two. Group One met at 8:00 in the morning and Group Two met at 10:00 the same morning.

FINDINGS

Responses to the preliminary test for knowledge of the meanings of the unusual English words are indicated in Table I. While the test

TABLE I.—NUMBER OF CORRECT AND ATTEMPTED DEFINITIONS OF THE UNFAMILIAR ENGLISH WORDS

	Number of subjects	Possible definitions per person	Correct definitions per person	Attempted definitions per person
Group One	63	24	0.40	0.83
Group Two	58	24	.67	1.40

results show that the words were not totally unfamiliar, the number of definitions attempted and, more particularly, the number of correct definitions, were so small that the use of the words as unfamiliar appeared to be justified.

In the subsequent discussion the familiar English words will be referred to as *X* words and the unfamiliar words as *Y* words.

Table II shows the mean errors for the two types of words on each of the three learning tests, the differences between the means, and the percentages realized of all the gain that was possible.

TABLE II.—AVERAGE NUMBERS OF ERRORS ON THE THREE IMMEDIATE TESTS AND PERCENTAGES OF POSSIBLE GAIN

Test	Mean errors type *X*	Mean errors type *Y*	Difference	Percentage of possible gain
Group One.				
I	17.25	.....	....	.28
II	8.05	.....	9.20	.53
III	3.92	.....	4.13	.51
I	.....	23.51	....	.02
II	.....	19.35	4.16	.18
III	.....	15.13	4.22	.22
Group Two.				
I	15.09	.....	....	.37
II	5.79	.....	9.30	.62
III	3.21	.....	2.58	.45
I	.....	22.91	....	.05
II	.....	17.86	5.05	.22
III	.....	12.41	5.45	.31

For both groups of subjects the *X* words carried a greater rate of learning between the first and second tests than did the *Y* words. The differences between the second and third testings were about the same for both types of words in Group One, but appeared to favor the *Y* words in Group Two. This indicated equality or even superiority of increase for the *Y* words, however, is a spurious one, since for both groups the number of unlearned *X* words has been reduced to less than half the number of unlearned *Y* words. The last column in the table shows clearly that with every presentation the percentage of possible gain realized for the *X* words surpassed the corresponding percentage of possible gain realized for the *Y* words.

Table III shows the number of words per person learned on each of the tests.

The difference per person for both groups favors words of the *X* type. Since the time allotment was constant for all, it follows that associations between familiar English words and Hindustani equivalents are built up more rapidly than associations between unfamiliar English words and Hindustani equivalents.

TABLE III.—LEARNING ACHIEVEMENT PER PERSON AS INDICATED BY THE LEARNING TESTS

Test	Total possible words	Words learned per person	Test	Total possible words	Words learned per person	Words per person difference ($X - Y$)
Group One.						
X I.....	24	6.75	*Y* I.....	24	0.49	6.26
X II.....	24	15.95	*Y* II.....	24	4.65	11.30
X III.....	24	20.08	*Y* III.....	24	8.87	11.21
Group Two.						
X I.....	24	8.91	*Y* I.....	24	1.09	7.82
X II.....	24	18.21	*Y* II.....	24	6.14	12.07
X III.....	24	20.79	*Y* III.....	24	11.59	9.20

TABLE IV.—A STUDY OF THE ERRORS ON THE THREE LEARNING TESTS FOR BOTH TYPES OF WORDS

Test	Mean	Standard deviation	Critical ratio	r	Probable error
Group One.					
X I.....	17.25	4.18	12.44	.30	.08
Y I.....	23.51	1.55			
X II.....	8.05	5.26	17.24	.43	.07
Y II.....	19.35	4.40			
X III.....	3.92	4.46	15.26	.42	.07
Y III.....	15.13	6.07			
Group Two.					
X I.....	15.09	4.45	12.80	.11	.09
Y I.....	22.91	1.94			
X II.....	5.79	5.11	17.50	.40	.17
Y II.....	17.86	4.44			
X III.....	3.21	4.03	11.03	.33	.08
Y III.....	12.41	6.40			

Certain additional information regarding learning appears when the gross data are refined.

Table IV brings out again the fact that fewer errors were made by both groups on words of the *X* type. Critical ratios indicate that the differences in learning are significant. Standard deviations indicate more homogeneity of learning for words of the *Y* type in the first and second testings, but the opposite is true on the third testing. This favoring of the *X* type on the third test is probably due, in part at least, to the fact that the spread had become limited. The correlations are low and probable errors do not compensate: It is possible that differences in the two kinds of learning are sufficiently pronounced so that ability in one is not a basis for predicting ability in the other.

In addition to rate and amount of growth, retention after two weeks was considered. Table V shows the relationship existing between *X* and *Y* words on the follow-up test.

TABLE V.—A RECORD OF THE MEANS, CRITICAL RATIO, STANDARD DEVIATIONS, CORRELATION, AND PROBABLE ERROR FOR THE FOLLOW-UP FOR THE TWO TYPES

Test	*M*	SD	CR	*r*	PE
Group One.					
X I *F*	12.02	5.24	16.05	.38	.07
Y I *F*	21.98	2.95			
Group Two.					
X I *F*	10.41	5.40	15.96	.40	.07
Y I *F*	21.00	3.17			

The mean error for type *X* words is significantly smaller than the mean error for type *Y* words. The standard deviation shows a wider dispersion of retention for *X* words than *Y* words. The correlation is not statistically significant.

Table VI gives the mean scores for the two types of words on the third learning test and the follow-up and shows the percentages of possible gain which were realized.

The table indicates that for Group One a somewhat larger number of *X* words than *Y* words was forgotten. But, since more than twice as many *X* words as *Y* words were learned for immediate recall, it is obvious that more *X* words than *Y* words were retained. Or, in other

TABLE VI.—AVERAGE NUMBERS OF ERRORS ON THIRD IMMEDIATE TEST AND FOLLOW-UP TEST AND PERCENTAGES OF RETENTION

Test	Mean errors type *X*	Mean errors type *Y*	Difference	Percentage of possible retention
Group One.				
III.....	3.92			
IV (follow-up).....	12.02	.....	8.10	.60
III.....	.....	15.13		
IV (follow-up).....	.....	21.98	6.85	.23
Group Two.				
III.....	3.21			
IV (follow-up).....	10.41	.....	7.20	.65
III.....	.....	12.41		
IV (follow-up).....	.....	21.00	8.59	.26

words, after two weeks sixty per cent of the *X* associations were recalled, as against twenty-three per cent of the *Y* associations.

In Group Two not only was the percentage of possible retention more than twice as great for the *X* words as for the *Y* words, but the average number of *X* words forgotten was less than the average number of *Y* words forgotten.

Table VII presents certain additional data.

TABLE VII.—A RECORD OF THE MEANS, CRITICAL RATIO, STANDARD DEVIATIONS, CORRELATION, AND PROBABLE ERROR FOR THE DIFFERENCES BETWEEN THE SCORES ON THE FINAL TESTING AND THE FOLLOW-UP IN THE SAME TYPE

Difference	*M*	SD	CR	*r*	PE
Group One.					
X III-*X* I *F*.....	8.10	4.38	1.76	.09	.08
Y III-*Y* I *F*.....	6.85	5.32			
Group Two.					
X III-*X* I *F*.....	7.21	4.68	−.127	−.02	.09
Y III-*Y* I *F*.....	8.59	5.17			

The average of the differences between the scores for the final testing and follow-up for one type of word is not significantly different from that for the other type, as shown by the low critical ratio. Standard deviations indicate slightly more homogeneity of retention for the *X* type than for the *Y* type. The correlation is low and the probable error high, suggesting that the two types of learning are not identical.

CONCLUSIONS

In the light of these data from two different groups of subjects using comparable but not identical material, it would appear that associations between English words and foreign language words are more readily formed when the English word is familiar to the learner. Not only does learning take place more quickly when the English equivalent is familiar, but the association is more permanent.

ABILITIES AS SUMS OF FACTORS, OR AS THEIR PRODUCTS

C. SPEARMAN

University of London

The theory of Two Factors, as is generally known, has its basis in the proposition that, when each of a set of variables can be expressed as the sum of two factors, the one being common to all the variables whilst the other is independent from one variable to another, then all the tetrads become zero.

But on what grounds, it may be asked, should the variable be just the *sum* of two factors? That it should be some or other function of them seems plausible enough. But the additive relation seems very unplausible.

An objection on this score might be expected from, and indeed has actually been raised by, that "Gestalt" school whose chief mission appears to be to charge "traditional" psychologists with building up their doctrine on additive lines. Now, the present writer for his part has, in spite of much searching, failed to discover any writer who really did build up his psychology on this notion of "Sums." But here, in this case of the factor theory, the offense has really and truly been committed. The factorist is caught *in flagranti delictu*. What about it?

The excuse, as all mathematicians and physicists know, is that the additive relation is the simplest possible one, that it is by far the easiest to manipulate mathematically, and above all things that in general it supplies an extraordinarily convenient first approximation. It is amazing how much of applied mathematics depends on the use of Taylor's theorem, which expresses the variation of any function of any number of elements as being approximately the sum of the (weighted) variations of the elements. In fact, the very basis of correlational theory involves the employment of this device.

But in the case of the proposition of the theory of two factors, there appears to be a new snag. In order really to support the theory, the proposition should only be found true when all approximations or other disturbances have been duly taken into account, *not before*. Too exact compliance of observation with the proposition would be nearly as bad as too inexact.

Now, as matters actually stand, the compliance seems to have gone about to the limit. Take, for example, the research of Stephenson with

over a thousand girls. When all due allowances had been made, the fulfilment of the proposition was so exact that no room appears to be left for any further disturbance at all, by approximations or otherwise. [He arrived at a median tetrad of 0.0081 with a PE of 0.0095.] In particular, there is no room for any appreciable disturbance due to using the sums instead of the true functions.

In these circumstances, it appeared interesting to ascertain how large any such disturbance really would be. For this purpose, the additive proposition might be compared with the far more plausible multiplicative one.

Now, the required comparison can be made as follows:

I. THE CASE OF ADDITION

In this case we have as usual $M_{ax} = C_a G_x + D_a S_{ax}$ (1)
where M_{ax} is the observed measurement of the ability a in the individual x. G_x and S_{ax} are the hypothetical general and specific factors, whilst C_a and D_a are constants for the ability a.

Then indicating the mean for all individuals by scoring, we get

$$\overline{M}_a + m_{ax} = C_a[\overline{G} + g_x] + D_a[\overline{S}_a + s_{ax}]$$

or, deducting the constants on each side,

$$m_{ax} = C_a g_x + D_a s_{ax} \tag{2}$$

whence

$$m_{ax} m_{bx} = C_a C_b g_x^2 + C_a D_b g_x s_{bx} + C_b D_a g_x s_{ax} + D_a D_b s_{ax} s_{bx} \tag{3}$$

whence, summing for all individuals, dividing by their number and taking all sigmas = unity, we arrive at

$$r_{ab} = C_a C_b,$$

whence by analogy

$$\frac{r_{ap}}{r_{aq}} = \frac{C_a C_p}{C_a C_q} = \frac{C_b C_p}{C_b C_q} = \frac{r_{bp}}{r_{bq}},$$

the known fundamental tetrad equation.

II. THE CASE OF MULTIPLICATION

In this second case the assumption becomes of the form

$$M_{ax} = E_a G_x S_{ax} \tag{4}$$

$\overline{M}_a + m_{ax} = E_a[\overline{G} + g_x][\overline{S}_a + s_{ax}]$, or deducting the constant constituent,

$$m_{ax} = E_a[\overline{S}_a g_x + \overline{G} s_{ax} + g_x s_{ax}], \tag{5}$$

which has the same form as in the case of Sums together with the further term $g_x s_{ax}'$. Next,

$$m_{ax} m_{bx} = E_a E_b[(\overline{S}_a g_x + \overline{G} s_{ax})(\overline{S}_b g_x + \overline{G} s_{bx}) + H], \tag{6}$$

where H contains the terms derived from the said additions and therefore amounts to

$$\overline{S}_a g_x{}^2 s_{bx} + \overline{G} g_x s_{ax} s_{bx} + \overline{S}_b g_x{}^2 s_{ax} + \overline{G} g_x s_{ax} s_{bx} + g_x{}^2 s_{ax} s_{bx}$$

If now we treat (6) as we did (3), all the terms from H vanish on using the well-known theorem of Isserlis.[1] There remains only

$$r_{ab} = E_a \overline{S}_a E_b \overline{S}_b, \tag{7}$$

whence just as before

$$\frac{r_{ap}}{r_{aq}} = \frac{r_{bp}}{r_{bq}}$$

Accordingly, it would seem that in the present case, the additive and the multiplicative functions lead to precisely the same proposition. If Stephenson's observed results leave no room for further allowance on account of employing sums instead of products, this fact completely agrees with our present finding that no such room is left theoretically. Thus, the congruence between observation and theory, far from being damaged by taking into consideration products in place of sums, is only verified the more.

---

[1] It assumes, as will be remembered, that the correlated variables are distributed normally. But this is a condition that in general can be satisfied at will (as was done by Stephenson himself).

GUESSING AND TELLING METHODS IN LEARNING WORDS OF A FOREIGN LANGUAGE

GEORGE FORLANO

The College of the City of New York

AND

MOSES N. H. HOFFMAN

Beth Jacob Hebrew School

When learning a foreign language, which is the better method to employ in acquiring the meaning of new words? Should the learner try to guess the meaning of a new foreign word by association, analysis, and comparison with subsequent active recall, or should he immediately be told the correct meaning? In other words, should the learner actively learn the new words, thereby usually making wrong responses before mastering the correct meaning; or should the learner acquire the correct meaning passively, that is, by being told, and thereby avoid making any wrong responses?

An attempt to throw light on this problem was made by setting up an experiment where these two methods of learning the meaning of new words; namely, the Guessing and the Telling Methods, would be compared as to efficient retention of words. These two methods of learning, although not representative of progressive practice among foreign-language teachers, have been studied in order to secure information as to whether the active phase in the Guessing Method will result in a greater positive influence for words learned than the mere associational phase as incorporated in the Telling Method.

The subjects employed were sixty-five pupils of a supplementary Hebrew school. Only girls attend the school. The pupils ranged in their public-school grades from the third through the eighth, and in chronological age from eight to fourteen years. The mean chronological age was eleven years and one month.

Mimeographed lists of Hebrew words, none of whose meanings was known to the subjects of the experiment, were the materials used. The English meanings, however, were all simple words known to the pupils. For example, the English meanings of the corresponding seven Hebrew words are: Silk, peanut, yawn, stretch, ugly, carrot, and zinc.

In order to equalize differences of such factors as age, previous training, and ability between the groups, difficulty of the lists of words, and practice effect, the experimental groups were rotated. Four

groups, each consisting of about sixteen girls, constituted the experimental groups. Two lists of twenty words each served as the material to be learned. The set-up of the experiment is shown in Table I. One group was taught the first list of twenty words using the Guessing Method, and then the second list of twenty words was taught according to the Telling Method; whereas another group learned the second list of twenty words by the Guessing Method, and then the first list by the Telling Method. Again, to a third group the first list was administered by the Telling Method first, and then the second list by the Guessing Method; while the fourth group was given the second list to learn by the Telling Method, and then the first list by the Guessing Method. The time taken in each experimental session was about an hour.

TABLE I.—SET-UP OF EXPERIMENT[1]

Lesson	Experimental groups			
	I	II	III	IV
1	Ga	Gb	Ta	Tb
2	Tb	Ta	Gb	Ga

[1] *G* = Guessing Method; *T* = Telling Method; *a*, *b* = lists of twenty word-pairs each.

The procedure was as follows: Blank cards to be used for covering the words were distributed to the children who were cautioned that "at no time should you write or mark anything upon the card." Then the experimenter said, "Today, I want to find out how many words you can learn. These are all new words, and I hope you will do your best to learn them. I shall now give each of you three lesson sheets. Leave them face down until I tell you what to do." When the pupils had filled in the blanks concerning the identity of the group, method of learning, name, age, grade, and date on each of the three learning sheets, the experimenter continued, "Now, take the first lesson sheet. I am going to show you how to learn these words. Listen carefully and follow the directions exactly. You will be given a test on these words after you have finished learning them.

"Now take the card and cover all the words except the first word. The word is _____." (The experimenter pronounces the Hebrew word.) For the Guessing Method the instructions read were as follows: "Now write next to it its meaning in English. If you don't know, guess. Write any word that comes to your mind first. Do not leave

any blank. Go ahead. (Pause.) I am now going to tell you the correct meaning. It is _____. If you have written the wrong meaning draw a line through it, and write the correct meaning next to it. If you had the correct meaning, write it again. I shall spell it for you: '_____.'

"Now move the card down so that you see the next word. No. _____. It is _____. (Experimenter pronounces the word.) Write its meaning next to it. If you don't know write any word that occurs to you. Write the meaning as soon as I say the Hebrew word. Remember, do not leave any blank. Go ahead. (Pause.) The correct meaning is _____. Draw a line through the wrong meaning and, next to it, write the correct one: _____." This procedure was employed for the other words in the list. The list of words was gone over three times. Each time a separate lesson sheet was used for learning.

For the Telling Method the directions were: "The word is _____. It means _____. Write the meaning next to it _____. Now move the card down so that you see the second word. It is _____. It means _____. Write the meaning next to it _____. Move the card down to No. _____: _____. (Here the experimenter read the number and the corresponding word.) It means _____." The same procedure was employed for the rest of the list. Here, too, the list of words was gone over thrice, each time using a separate lesson sheet. The learning time was the same for either method, namely, eighteen minutes.

Immediately after each learning period an immediate retention test was given, for which the directions were as follows: "Write the English meaning next to the Hebrew word. The words here are *not* in the same order as on the learning sheets. Do not spend too much time on any one word. First answer those which you know, and then, if there is time left, go back to those you have omitted."

In order to determine the relative effectiveness of the two methods for delayed retention, the pupils were given a retention test again after a period of two days.

The results of the experiment are presented in Table II. It is seen that on the immediate retention test the mean for the Telling Method surpasses the mean for the Guessing Method by more than three words. On the delayed retention test using the recall type of test, the Telling Method is superior by almost two words. All the mean differences are statistically reliable. The standard error of the mean difference was obtained in each case by the use of the long formula,

$$\sigma_d = \sqrt{\sigma^2_{M_1} + \sigma^2_{M_2} - 2r\sigma_{M_1}\sigma_{M_2}}.$$

TABLE II.—RELIABILITY OF DIFFERENCES BETWEEN MEANS OF THE GUESSING AND TELLING PERFORMANCES ON THE IMMEDIATE AND DELAYED RETENTION TESTS

Test	Learning methods						Mean Diff.	σ_d	Standard Ratio[1]
	Guessing			Telling					
	Range	Mean	σ	Range	Mean	σ			
Immediate.....	0–14	4.66	3.42	0–20	8.10	5.50	3.44	.50	6.88
Delayed (Recall).....	0–11	3.92	3.22	0–20	5.80	4.38	1.88	.48	3.92

[1] The standard ratio is the ratio of the mean difference to its standard error.

We may conclude from our results that in teaching the meanings of a list of words of a foreign language it is better to tell the correct meaning immediately without allowing the learner to guess a possible wrong meaning. It must be borne in mind that the present experiment was concerned with learning the meaning of a list of isolated words, not new words encountered in context. Theoretically, the experimental groups were rotated so as to equalize all differences save the important difference between the Telling Method and the Guessing Method; namely, that if an error is to be made it could only occur in the latter method. We may conclude that the errors made during the learning by the Guessing Method hampered learning.

This is a conclusion that is not unexpected since the work of Lorge (1), Lorge and Thorndike (2), Thorndike (3), and Tuckman (4) have demonstrated that an error having occurred does more harm by being permitted than punishment can offset. It is obvious that the punishment in our experiment is the knowledge of a wrong guess. There seems to be implied in these laboratory studies that the sheer load of vocabulary requisite for the learning of language may be acquired more efficiently by direct associational learning than by learning conceived as an active process on the part of the individual where such activity partakes of the nature of random rather than directed participation.

REFERENCES

1. Lorge, Irving: "The Efficacy of Intensified Reward and of Intensified Punishment." *Journal Experimental Psychology*, Vol. XVI, 1933, pp. 177–207.

2. Lorge, Irving and Thorndike, Edward L.: "The Comparative Strengthening of a Connection by One or More Occurrences of It in Cases Where the Correction Was Punished and Was Neither Punished nor Rewarded." *Journal of Experimental Psychology*, Vol. XVI, 1933, pp. 374–382.
3. Thorndike, Edward L.: *Fundamentals of Learning.* New York: Teachers College Bureau of Publications, 1932, p. 291.
4. Tuckman, Jacob: *The Influence of Varying Amounts of Punishment on Mental Connections.* Contributions to Education, No. 590, Teachers College, Columbia University, 1933.

COMPARISON OF SCALES FOR EVALUATION OF PERIODICALS

WALTER CROSBY EELLS

Coördinator, Coöperative Study of Secondary School Standards, 744 Jackson Place, Washington, D. C.

In an earlier issue of the *Journal of Educational Psychology* appears an account of the derivation of a scale for measuring the cultural content of general interest magazines.[1] It was based upon the judgments of sixty-five individuals who were asked to rate a selected list of periodicals on their general cultural content, cultural being defined as "pertaining to the degree of refinement of mind, morals, or tastes." The resultant scale involved weights from ten to one, the limits of the ten classes being determined by half-sigma intervals of the entire distribution, for the seventy-four magazines evaluated.

Using a closely analogous process the Coöperative Study of Secondary School Standards[2] has derived a scale for the evaluation of the periodicals in the libraries of secondary schools. The Coöperative Study Periodical Scale is based upon the judgments of one hundred sixty carefully selected librarians of various types who were asked to evaluate a list of one hundred periodicals, commonly found in secondary school libraries, on their value "for general purposes for the typical high-school library."[3]

---

[1] Morgan, Winona L., and Leahy, Alice M.: "The Cultural Content of General Interest Magazines." *Journal of Educational Psychology*, Vol. XXV, October, 1934, pp. 530–536.

[2] The Coöperative Study of Secondary School Standards is a joint undertaking of the six regional associations of colleges and secondary schools of the country—the New England, Middle States, North Central, Southern, Northwest, and Western Associations, which for three years has been studying secondary schools with a view to the improvement of evaluation and accreditation procedures. During 1936–1937 it studied intensively secondary schools in every state chosen so as to represent, on a proportional basis, schools of all types, public and private, large and small, urban and rural, coeducational and for one sex only, progressive and conservative. The total number of schools was two hundred. Special attention was devoted to the libraries of these schools, their book collections, their periodicals, and student and faculty use of library resources. The scale described here was developed in order to have a more valid measure of the quality and quantity of the periodicals found in the school libraries.

[3] For details of the derivation of this scale, and a list of the one hundred and twenty-seven magazines evaluated, with the weight given to each, see Eells, Walter Crosby: "Scale for the Evaluation of Periodicals in Secondary-School Libraries," *Wilson Bulletin for Librarians*, Vol. XI, June, 1937, pp. 668–673.

TABLE I.—COMPARISON OF CLASS VALUES FOR THIRTY-NINE PERIODICALS ON THE COÖPERATIVE STUDY AND GENERAL CULTURAL SCALES

Title	Weights on Coöperative Study scale	Weights on general cultural scale	Difference Coöperative Study—general cultural
National Geographic Magazine	10	8	+2
Current History Magazine	9	8	+1
Literary Digest	9	7	+2
Popular Mechanics Magazine	9	5	+4
Popular Science Monthly	9	5	+4
Readers Digest	9	7	+2
Harpers Magazine	8	9	−1
Atlantic Monthly	8	10	−2
Hygeia	8	6	+2
Scientific American	8	7	+1
Review of Reviews and World's Work	8	7	+1
Time	7	7	0
Survey	7	8	−1
Scribners Magazine	7	8	−1
Saturday Review of Literature	7	10	−3
Travel	6	7	−1
Good Housekeeping	6	6	0
Forum and Century	6	9	−3
New Republic	6	9	−3
Nation	6	9	−3
Asia	6	8	−2
Better Homes and Gardens	6	6	0
House Beautiful	5	7	−2
House and Garden	5	6	−1
Fortune	5	7	−2
Nations Business	4	6	−2
Living Age	4	8	−4
Field and Stream	4	6	−2
Parents Magazine	4	6	−2
American Magazine	4	5	−1
Saturday Evening Post	3	5	−2
Womans Home Companion	3	5	−2
Ladies Home Journal	3	5	−2
Delineator	3	5	−2
Pathfinder	2	5	−3
McCall's Magazine	2	4	−2
Country Gentleman	2	6	−4
American Mercury	2	8	−6
Colliers	1	5	−4
Average	5.7	6.8	1.1

COMPARISON OF THE TWO SCALES

Value from the standpoint of "general cultural content" may be quite different in some cases from "general purposes for the typical secondary-school library," but there should be many points of similarity. The Morgan-Leahy scale contained seventy-four titles, the Coöperative Study Scale contains one hundred twenty-seven titles. Table I shows the thirty-nine titles common to the two scales, with the relative class weights on each scale. From the methods of derivation of these two scales, the two columns of weights are directly comparable; a weight of six, for example, on either scale, including the periodicals found between the means and points one-half standard deviation above the means of the judgments of the two groups of judges.

On the average, the cultural scale places this particular group of magazines slightly more than one class higher than does the secondary-school scale. In the case of nine titles the secondary-school scale rates them higher than the cultural scale, in three cases there is no change, while in twenty-seven cases the cultural scale ranks them higher. In only six cases, however, is the difference more than three, the most outstanding one being *American Mercury* which is given a weight of eight from the standpoint of general culture but of only two for secondary-school purposes. Most of the differences noted can be understood and justified in view of the different criteria used in the construction of the two scales.

COMPARISON WITH C.C.C. LIST

It is of interest, also, to compare the secondary-school scale with the standard list of forty-four magazines furnished by the United States Army for the C.C.C. camps.[1] Only sixteen titles are common to the two lists, as shown in Table II.

The average of these weights, is 0.6 lower than the average of the listed magazines which the Coöperative Study has found in two hundred secondary-school libraries which it has evaluated, this average being 6.9. This applies, however, to only sixteen of the forty-four magazines on the C.C.C. list. If a score of 1 is given for each of the others, as was done in the use of the Coöperative Study Scale, the average drops to only 1.7, distinctly below the poorest school library collection studied.

---

[1] Tucker, William P.: "C.C.C. Library Work," *School and Society*, Vol. XLIII, January 4, 1936, p. 20.

TABLE II.—RATINGS OF C.C.C. MAGAZINES ON COÖPERATIVE STUDY PERIODICAL SCALE

Title	Weight on coöperative study scale	Rank on coöperative study scale
National Geographic Magazine	10	2
Current History Magazine	9	5
Literary Digest	9	6
Nature Magazine	9	7
Popular Mechanics Magazine	9	8
Popular Science Monthly	9	10
Readers Digest	9	11
Time	7	24
Radio News and the Short Wave	6	47
House Beautiful	5	84
Field and Stream	4	104
American Magazine	4	107
Leisure	4	108
Saturday Evening Post	3	111
Mid-week Pictorial	3	115
Colliers	1	127

Of the C.C.C. list, twenty-five are found on the general cultural scale, with an average class value of only 4.4, which drops to 2.9 if a value of 1 is assigned to the titles not on the cultural scale. Possibly the camps are justified in stressing recreational reading rather than cultural, but the two are not necessarily inconsistent. These data suggest the possibility of improvement of the "standard list" furnished by the Army, especially when one notes that two copies each of *Colliers* (Value 1), *Liberty* (Value 1), and the *Saturday Evening Post* (Value 3) are furnished, but only one of the *Readers Digest* (Value 9).